# RESEARCH IN THE SPACE SCIENCES

A Bibliography of Research Performed
Through the Space Sciences Laboratory,
University of California, Berkeley
with partial support from
NASA Grant NGL 05-003-497

Kinsey A. Anderson, Director
and
Principal Investigator

Volume I, No. 1

January - March, 1974

(NASA-CR-155674) RESEARCH IN THE SPACE SCIENCES, VOLUME 1, NO. 1 Progress Report, Jan. - Mar. 1974 (California Univ.) 297 p
00/99
N78-73055 THRU N78-73067
Unclas
04798

Space Sciences Laboratory, Series 15, Issue 60

REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

TABLE OF CONTENTS

L.W. Alvarez and Group

Electronic Processes in Liquid Xenon, S.E. Derenzo, T.S. Mast, H. Zaklad and R.A. Muller, Nucl. Inst. & Methods, 118, 611, 1974

Electron Avalance in Liquid Xenon, S.E. Derenzo, T.S. Mast, Haim Zaklad and R.A. Muller, Phys. Rev. A9, 2582, 1974

Searching for Antimatter in the Cosmic Rays, A. Buffington, Atmosph. Tech. 5, 25, 1974

K.A. Anderson and Group

30 to 100 keV Protons Upstream from the Earth's Bow Shock, R.P. Lin, C.-I. Meng and K.A. Anderson, J. Geophys. Res., 79(4), 489, 1974

Suprathermal Particles, R.P. Lin, Invited review paper, presented at the Third Solar Wind Conference, Asilomar, Pacific Grove, California, March 1974

Magnetic Field Configuration in the Magnetotail near 60 $R_e$, C.-I. Meng and K.A. Anderson, March 1974

C.S. Bowyer and Group

On the Distribution of $He^+$ in the Plasmasphere from Observations of Resonantly Scattered HeII 304-Å Radiation, F. Paresce, C.S. Bowyer and S. Kumar, J. Geophys. Res., 79, 174, 1974

Observations of HeI 584Å Nighttime Radiation: Evidence for an Interstellar Source of Neutral Helium, F. Paresce, S. Bowyer and S. Kumar, Astrophys. J., 187, No. 3, part 1, 1974

On the Opacity of the Interstellar Medium to Ultrasoft X-rays and Extreme Ultraviolet Radiation, R. Cruddace, F. Paresce, S. Bowyer and M. Lampton, Astrophys. J., 187, No. 3, part 1, 1974

Cosmic Far Ultraviolet Background, A. Davidsen, S. Bowyer and M. Lampton, Nature, 247, No. 5442, 513, 1974

Further Evidence for an Interstellar Source of Nighttime HeI 584Å Radiation, F. Paresce, S. Bowyer and S. Kumar, Astrophys. J., 188, L71, 1974

Extreme Ultraviolet Spectrometer for Space Research, S. Kumar, F. Paresce, S. Bowyer and M. Lampton, Applied Optics, 13, 575, 1974

A.L. Burlingame and Group

The Ion Implantation Chemistry of Carbon in Metals and Mineral Grains and its Implications for Cosmochemistry, J.P. Bibring, A.L. Burlingame, J. Chaumont, M. Maurette, F.C. Walls, P.C. Wszolek, Trans. Am. Geophys. Union 54(11), 1129, 1973

Meteoritic Contribution to Lunar Carbon Chemistry, P.C. Wszolek, B.R. Simoneit, P.T. Holland and A.L. Burlingame, Trans. Am. Geophys. Union 54, 355, 1973

Mechanisms of Unimolecular Gas-Phase Gamma-Hydrogen Rearrangements of Radical Cations. The Influence of Entropy as Opposed to Energy, P.J. Derrick, A.M. Falick and A.L. Burlingame, J. Am. Chem. Soc., 96, 615, 1974

Progressive Specific Hydrogen Rearrangements Quenched by Molecular Ion Fragmentation in 2-Methylpropene. Mass Spectral Serendipity, P.J. Derrick and A.L. Burlingame, J. Am. Chem. Soc., 96, 4909, 1974

Deuterium and Carbon-13 Tracer Studies of Ethanol Metabolism in the Rat by $^{2}H$, $^{1}H$-Decoupled $^{13}C$ Nuclear Magnetic Resonance, D.M. Wilson, A.L. Burlingame, T. Cronholm and J. Sjovall, Biochem. and Biophysical Res. Comm., 56, 828, 1974

Hydrogen-Deuterium Randomization in 7-Methyl-4-octanone-7-$d_1$ at Times of $10^{-11}$-$10^{-5}$ Sec Following Field Ionization, P.J. Derrick, A.M. Falick, A.L. Burlingame, and C. Djerassi, J. Am. Chem. Soc., 96, 1054, 1974

Apparent Charge Localization Effects in the McLafferty Rearrangement of Acetate Esters: Field Ionization Kinetic Measurements, G. Wood, A.M. Falick and A.L. Burlingame, Organic Mass Spec., 8, 279, 1974

C. W. Churchman and Group

Systems Analysis as Technology Transfer, I.R. Hoos, J. Dynamic Systems, Meas. and Control, March 1974

B. Fraenkel-Conrat

Correlation Between Amino Acid Exchanges in Coat Protein of TMV Mutants and the Nature of the Mutagens, B. Singer and H. Fraenkel-Conrat, Virology, 60, 485, 1974

H. Heckman and Group

A Facility for Calibrations and Short Experiments with Heavy Ions at the Bevatron, D.E. Greiner, P.J. Lindstrom, F.S. Bieser and H.H. Heckman, Nucl. Instru. and Methods, 116, 21, 1974

H.O. Kammen

Biosynthesis of Pseudouridine in Transfer Ribonucleic Acid, R. Cortese, H.O. Kammen, S.J. Spengler and B.N. Ames, J. Biolog. Chem., 249, 1103, 1974

T.H. Jukes and Group

On the Possible Origin and Evolution of the Genetic Code, T.H. Jukes, Origins of Life, 5, 331-350, 1974

Conservation of Shannon's Redundancy for Proteins, L.L. Gatlin, J. Mol. Evol., 3, 189-208, 1974

M.F. Maestre

Circular Dichroism of Films of Polynecleotides, W.C. Brunner and M.F. Maestre, Biopolymers, 13, 345-357, 1974

Fluorescence Detected Circular Dichroism, D.H. Turner, I. Tinoco, Jr., M.F. Maestre, J. Am. Chem. Soc., 96, 4340, 1974

Mie Scattering Contributions to the Optical Density and Circular Dichroism of T2 Bacteriophage, G. Holzwarth, D.G. Gordon, J.E. McGinness, B.P. Dorman, M.F. Maestre, Biochemistry, 13, 126, 1974

F.S. Mozer and Group

The Impact of Space Electric Field Research on Atmospheric Studies, F.S. Mozer, Acta Astronautica, 1, 71-79, 1974

The Average Auroral Zone Electric Field, F.S. Mozer and P. Lucht, J. Geophys. Res. 79(7), 1001, 1974

P.B. Price and Group

Gas-Rich Meteorites: Possible Evidence for Origin on a Regolith, D. MacDougall, R.S. Ragan, P.B. Price, Science, 183, 73-74, 1974

Irradiation Records in Orange Glass and Two Boulders from Apollo 17, D. MacDougall, I.D. Hutcheon, P.B. Price, Proc. Fifth Lunar Sci. Conf., 1974

Analytic Variational Calculation of the Ground-State Binding Energy of Hydrogen in Intermediate and Intense Magnetic Fields, L.W. Wilson, Astrophys. J., 188, 349-352, 1974

C.H. Townes and Group

The Methanol Source in Orion at 1.2 Centimeters, M.F. Chui, A.C. Cheung, D. Matsakis, C.H. Townes, A.G. Cardiasmenos, Astrophys. J., 187, L19-L21, 1974

Detection of $^{17}O$ in IRC+10216, D.M. Rank, T.R. Geballe, E.R. Wollman, Astrophys. J., 187, L111-L112, 1974

NUCLEAR INSTRUMENTS AND METHODS 118 (1974) 611–612; © NORTH-HOLLAND PUBLISHING CO.

# ELECTRONIC PROCESSES IN LIQUID XENON*

S. E. DERENZO, T. S. MAST and H. ZAKLAD

*Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, U.S.A.*

R. A. MULLER

*Space Science Laboratory, University of California, Berkeley, California 94720, U.S.A.*

Received 9 October 1973

This letter points out several basic errors that appeared in a recent article by Prunier et al. These errors relate to the analysis of data taken in liquid-xenon single-wire cylindrical chambers.

Several basic errors appeared in an article recently published by Prunier et al. entitled *Some properties of xenon liquid-filled nuclear detectors*[1]). The article describes an experiment to measure electronic phenomena in liquid xenon using single-wire cylindrical chambers. We describe below some errors made in their interpretation of their experimental observations.

1) Prunier et al.[1]) incorrectly analyzed their collected charge vs applied voltage data by assuming that the *observed* charge collected in their single-wire proportional chamber is equal to the *total* charge separated in the avalanche. In general these two charges are very different. For many years this difference has been known to exist in the analogous situation of gas-filled proportional chambers. In addition, our group in Berkeley has taken similar data and performed an analysis to determine the recombination of the initial ionization and the first Townsend coefficient as a function of electric field[2, 3]). A rigorous analysis[3]) shows that only approximately 10% of the avalanche charge is induced in the first 2 $\mu$s and that the avalanche electrons only contribute approximately 10%.

The following calculation will illustrate this point: At a typical electric field of $1.5 \times 10^6$ V/cm the mean free path for an ionization collision ($\lambda_c = 1/\alpha$) is $(0.35 \pm 0.05)$ $\mu$m, taken from ref. 3. We believe that value to be more reliable than the value of 0.5 $\mu$m reported in ref. 1 (fig. 6), but the choice does not materially affect the argument. The avalanche may be closely approximated as the abrupt separation of an electron charge $Q$ and an equal $Xe^+$ ion charge at an effective distance $\lambda_c$ from the surface of the wire (the location of the center of gravity of the avalanche). The electrons drift to the wire at their saturated drift velocity[4]) of $3 \times 10^5$ cm/s, in a time $10^{-10}$ s. During this time, a charge of only $0.012\,Q$ is induced at the anode[5]) (assuming an anode diam. of 7.5 $\mu$m and a cathode diameter of 10 mm). An even larger fraction of $Q$ is induced by the outward motion of the $Xe^+$ ions during the first 2 $\mu$s (a typical pulse-shaping time – the calculation is very insensitive to this value). During 2 $\mu$s, the $Xe^+$ ions move outward[6]) 5.1 $\mu$m and induce an additional pulse of $0.11\,Q$. The $Xe^+$ ion mobility[7]) is $3 \times 10^{-4}$ cm$^2$/Vs. The remaining charge ($0.88\,Q$) will be induced *after* the pulse has peaked, during the 0.74 s required for the $Xe^+$ ions to reach the outer electrode.

The authors do not mention their pulse rise time or the mobility of $Xe^+$ ions, essential factors for the interpretation of their data.

2) Prunier et al. ignore the recombination of the initial electrons and $Xe^+$ ions, a process that occurs very soon after the passage of the recoil electron. As determined in ref. 3, $((26 \pm 2)\%$ of the charge is lost in liquid xenon through recombination[8]) at 280 V/cm, the average field in the chamber of Prunier et al. at the applied potential of 500 V. This is in agreement with the "gain" of 0.75 shown in fig. 5 of their paper at 500 V, assuming an uncollimated source. We conclude that although the reduced pulse heights observed by Prunier et al. at low voltage could in part be due to impurities, nearly all of the effect is due to recombination.

3) In section 4.2, it is explained why break-down should occur at gains of 65–100, in agreement with observation. This agreement can only be due to cancelling errors, since in reality there are approximately *ten times* as many avalanche electrons (and $Xe^+$ ions) as Prunier et al. assumed. Our analysis[3]) shows that the secondary emission coefficient $\gamma$ is less than $3 \times 10^{-4}$, in contradiction to the claim of

* Work done under the auspices of the U.S. Atomic Energy Commission and the National Aeronautics and Space Administration.



REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

Prunier et al. that $\gamma$ lies between $10^{-2}$ and $1.5 \times 10^{-2}$.

4) There are several errors relating to the electron capture coefficient $\beta$. From the work of Swan[9]) and Miller et al.[4])

$$\beta/P_0 = A/E,$$

where $\beta$ is the capture probability per mm drift, $P_0$ is the oxygen concentration in liquid argon, $E$ is the electric field, and $A$ is a constant. The value $A = 7$ per (mm V cm$^{-1}$ ppm) given by Prunier et al. in section 4.1 is incorrect in both magnitude and units. The correct value is $A = 700$ V cm$^{-1}$ per (mm ppm) which is evident from the original data shown in fig. 7. Furthermore, all left-hand vertical scale markers in fig. 6 should be multiplied by 10 [so that $A = 500$ V cm$^{-1}$ per (mm ppm) for $O_2$ in liquid xenon]. We also point out that the horizontal scale in fig. 9 should read V/cm rather than kV/cm.

5) In section 1, paragraph 4, it is stated that (when sufficiently purified) the rare gas solids and liquids, and semiconductors are the only known condensed dielectrics in which electrons remain free. Prunier et al. are apparently unaware of research performed over many years with room-temperature hydrocarbons[10, 11]). Note that above 5 kV/cm, electrons travel faster[11]) in $Si(CH_3)_4$ than in any liquefied noble gas.

Although there are serious errors in their analysis, the raw experimental data of Prunier et al. are in substantial agreement with ours. For several years our group at Berkeley has been operating liquid-xenon-filled single-wire proportional chambers[2, 3]), multi-wire ionization chambers[12]), and multi-wire proportional chambers[13–15]). Recently, we have obtained images of distributed $\gamma$-ray sources, using lead collimators and a 24-wire liquid-xenon proportional chamber[15]). This rapidly expanding technology offers unique possibilities for studies of the electronic states of simple liquids as well as important applications in particle physics and medicine.

## References


1) J. Prunier, R. Allemand, M. Laval and G. Thomas, Nucl. Instr. and Meth. **109** (1973) 257.

2) R. A. Muller, S. E. Derenzo, G. Smadja, D. B. Smith, R. G. Smits, H. Zaklad and L. W. Alvarez, Phys. Rev. Letters **27** (1971) 532.

3) S. E. Derenzo, T. S. Mast, H. Zaklad and R. A. Muller, Lawrence Berkeley Laboratory Report LBL-1313 (1973), submitted to Phys. Rev. A.

4) L. S. Miller, S. Howe and W. E. Spear, Phys. Rev. **166** (1968) 871.

5) H. H. Staub, in *Experimental nuclear physics* (Vol. 1, ed. E. Segré; J. Wiley, New York, 1952) p. 16.

6) H. H. Staub, p. 35 of ref. 5.

7) H. Davis, S. Rice and L. Meyer, J. Chem. Phys. **37** (1962) 947.

8) Experiments have shown that the fraction of electrons escaping recombination can be represented by $(1+K/E)^{-1}$, where $E$ is the electric field, and $K = (1100 \pm 200)$V/cm for minimum ionization tracks in liquid argon [John Marshall, Rev. Sci. Instr. **25** (1954) 232] and $K = (98 \pm 12)$V/cm for minimum ionization tracks in liquid xenon (ref. 3).

9) D. W. Swan, Proc. Phys. Soc. **83** (1968) 659.

10) P. H. Tewari and G. R. Freeman, J. Chem. Phys. **49** (1968) 4394; and **51** (1969) 1276.

11) W. F. Schmidt and A. O. Allen, J. Chem. Phys. **50** (1969) 5037; and **52** (1970) 4788.

12) S. E. Derenzo, A. Schwemin, R. G. Smits, H. Zaklad and L. W. Alvarez, Lawrence Berkeley Laboratory Report LBL-1791, to be published in the Proc. 1973 Intern. Conf. on *Instrumentation for high-energy physics*, Frascati, Italy (May, 1973).

13) H. Zaklad, S. E. Derenzo, R. A. Muller, G. Smadja, R. G. Smits and L. W. Alvarez, IEEE Trans. Nucl. Sci. NS-**19** no. 3 (1972) 206.

14) H. Zaklad, S. E. Derenzo, R. A. Muller and R. G. Smits, IEEE Trans. Nucl. Sci. NS-**20**, no. 1 (1973) 429.

15) S. E. Derenzo, T. F. Budinger, R. G. Smits, H. Zaklad and L. W. Alvarez, Lawrence Berkeley Laboratory Report LBL-2092, to be published in the Proc. Symp. on *Advanced technology arising from particle physics research*, Argonne National Laboratory, Argonne, Illinois (May 17, 1973).

# ELECTRON AVALANCHE IN LIQUID XENON*

Stephen E. Derenzo, Terry S. Mast, and Haim Zaklad
Lawrence Berkeley Laboratory, University of California
Berkeley, California 94720
and
Richard A. Muller, Space Science Laboratory, University of California
Berkeley, California 94720

A74-33543

March 1973

ABSTRACT

We present detailed measurements of the electron avalanche process in liquid Xenon. The measurements were made by using liquid-Xe-filled proportional chambers with anode diameters of 2.9, 3.5, and 5.0 μ to detect 279-keV γ rays and measure the photopeak pulse height as a function of applied voltage. The use of uniform pulses of electrons enabled us to discriminate against secondary Townsend processes. We present a table of the first Townsend coefficient α as a function of electric field E; a typical value is $\alpha = (4.5\pm0.3) \times 10^4$ cm$^{-1}$ at $E = 2 \times 10^6$ V/cm. The electron avalanche occurs in liquid Xe at electric fields 26 times smaller than would be predicted using measurements made in gaseous Xe and E/ρ density scaling.

Physical Review, A9, 2582 (1974)

## I. INTRODUCTION

It has been known for over two decades that excess electrons (such as those freed by the passage of ionizing radiation) in very pure liquified Ar and Xe remain free and have high mobility in applied electric fields.[1-5] These mobility data have been successfully interpreted in terms of gas kinetic theory, implying that electrons move in these liquids as they would in compressed gases.[6,7]

In 1968 Alvarez suggested[8] that the noble liquids might have important applications for high-energy nuclear particle detection, particularly if the electronic signal could be increased by inducing an electron avalanche in the liquid (similar to that which occurs in gas-filled detectors). By using suitably high electric fields we were able to induce electron avalanche in both liquid Ar and liquid Xe.[9-11] Only in liquid Xe did we observe a "well behaved" avalanche (i.e., one which had characteristics similar to those of a gaseous avalanche).[12] In this paper we report measurements in liquid Xe of the characteristic function that describes the avalanche, the first Townsend coefficient.

Electron avalanche occurs when the drifting electrons attain sufficient kinetic energy from the electric field to cause additional ionization. The first Townsend coefficient α is defined by the equation

$$dQ = \alpha Q\, dr, \qquad (1.1)$$

where dQ is the charge of electrons (and $Xe^+$ ions) liberated in the avalanche when an electron charge Q drifts through a distance dr; α can be a function of electric field, density, temperature, pressure, and the other parameters of the gas or liquid. In our experiments the only such parameter that varied by a significant amount was the electric field E. We obtained data on the avalanche by studying the pulses in a liquid-

filled proportional chamber as a function of applied voltage and anode diameter. Preliminary results were reported previously.[10]

## II. DESCRIPTION OF THE EXPERIMENT

The experimental apparatus, a single-wire proportional chamber, is shown in Fig. 1. The chamber cathode was 8 mm in diameter and 30 mm long, formed by depositing a conductive coating of SnO on the cylindrical inside glass wall of the chamber. The anode was a fine tungsten wire, alternately 2.9, 3.5, or 5.0 μ in diameter.[13] The wires were tensioned at 60% of their breaking load and soldered to feed-throughs at the top and bottom of the chamber. Several samples of each wire were photographed with a scanning electron microscope that was calibrated with a Buckbee-Meers screen[14] having 2000 meshes per inch (see Fig. 2). Variations in wire diameter were typically ±1.5%. To drive contaminants from the chamber walls, the chamber was kept at a temperature of 100°C and an internal pressure below $10^{-6}$ Torr for a period of at least 15 hours prior to filling. The chamber was filled with Xe gas from a specially designed noble gas purifier[15] and then cooled to -107°C (the boiling point of liquid Xe) by immersion in a thermostatically controlled bath of Freon-11 ($FCCl_3$). Although our purifier reliably delivers Xe gas that contains <50 ppb of $O_2$ and <100 ppb of either $N_2$, $CO_2$, or $CH_4$, the chamber fills with a liquid that typically contains electronegative impurities (as yet unidentified) equivalent in electron capture ability to several ppm $O_2$. (See Ref. 15 for details.) Difficulty in removing electronegative impurities from liquid Xe has been mentioned by other workers.[16,17] We are able to eliminate almost all of these electronegative impurities by using field emission electrons to sweep them to the outer electrode. A potential of approximately 5 kV is applied to the outer electrode, causing



the fine central wire to emit (typically) 5μA of current. After 10 min of this electronegative ion pumping the pulse height reaches a level that is not increased by further processing. (It may be readily calculated that < 300 ppb of impurities are removed.) Re-processing is required only after a period of several hours. These techniques are capable of reducing the impurity level to the point where electrons have a mean free path >20cm, as measured in a parallel-plate ionization chamber with a Frisch grid.

A collimated $^{203}Hg$ source was used to send a 1.2-mm-diameter beam of 68-keV x-rays and 279 keV gamma rays into the 8-mm-diameter cylindrical chamber. The resulting recoil electron pulses were detected with a charge-sensitive amplifier of the type used for solid state detectors. Electronic filtering consisting of 1-μsec integration (risetime) and 5-μsec differentiation (fall time) was chosen to maximize the signal-to-noise ratio. With this filtering, the noise was equivalent to $1.2 \times 10^{-16}$C FWHM at the input of the charge-sensitive amplifier. The chamber capacitance was 4.5 pF. A typical pulse height distribution for the $^{203}Hg$ source (recorded on a RIDL model 34-12B pulse height analyzer) and shown in Fig. 3, clearly shows the photopeak and Compton edge for the 279-keV γ-ray and the photopeak for the 68-keV x ray. The pulse height of the 279 keV photopeak is shown in Fig. 4 as a function of applied voltage for 2.9, 3.5, and 5.0 μ diameter tungsten wires. The amplifiers and pulse-height analyzer were calibrated by injecting pulses of known voltage and 200-nsec risetime through a 10 pF capacitor directly into the charge sensitive amplifier.

## III. INITIAL ANALYSIS

The data in Fig. 4 have the following qualitative interpretation: At the lowest voltages an appreciable fraction of the electrons recombine with $Xe^+$ ions before the electrons can contribute to the pulse. As the voltage is increased nearly all the electrons escape recombination, and the pulse height reaches a plateau. The height of this plateau varies by about 10% from run to run due to variations in the level of electronegative impurities or possible calibration errors. At higher voltages, the electrons avalanche in the high electric field near the anode, resulting in a rapid increase of pulse height with voltage. At our highest voltages the space charge of the slow $Xe^+$ ions reduces the field near the wire, causing a saturation of pulse height and a broadening of the photopeak. (We found that the saturation charge decreased by a factor 3 when the counting rate was increased from 200 to 2000 $sec^{-1}$ by tripling the diameter of the collimator hole. The data in Fig. 4 were taken with a counting rate of 200 $sec^{-1}$).

At voltages slightly beyond those shown in Fig. 4, the photopeak is obscured by a large number of pulses (not related to the presence of the $\gamma$-ray source) of irregular shape and size. Investigation at higher voltages is prevented by the onset of sparking through the liquid at many places between the cathode and the anode wire. When the number of sparks exceeds $\sim 10^3$, the anode wire often breaks.

One estimate of the first Townsend coefficient may be obtained by comparing the pulse heights of two chambers having different-size wires but the same electric field function E(r) where r is the distance to the center of the wire. The ratio of these pulse heights will be the additional gain from the avalanche that occurs in the distance from the larger-wire radius to the smaller-wire radius.



For example, one obtains the same function E(r) (in the region $r \geq 1.75$ μ) for a 2.9-μ-diameter wire at 1800 volts as for a 3.5-μ-diameter wire at 1765 volts. From Fig. 4, we see that for the two wires at their respective voltages the pulse heights are 6.6 and 2.9 $\times 10^{-15}$C. As we expect the avalanche to be identical in the regions where the fields are identical, we conclude that the additional gain of 6.6/2.9 = 2.3 occurred during the last 0.3 μ of the avalanche, i.e., as the avalanche progressed from the virtual surface of the larger wire at r = 1.75 μ to the actual surface of the finer wire at r = 1.45 μ. Assuming that the ratio of the gains is equal to the ratio of the size of the avalanches, this implies an average first Townsend coefficient of $\alpha = \ln(2.3)/0.3\ \mu = 28\,000\ cm^{-1}$. The electric field in this region varied from $1.29 \times 10^6$ to $1.57 \times 10^6$ V/cm.

In order to remove the equal-ratio assumption and to study the behavior of the first Townsend coefficient in detail, we made a simultaneous fit to all the data of Fig. 4 below 200 fC. We now consider the details of that fit.

## IV. DETAILED ANALYSIS

TABLE I. Glossary

| | |
|---|---|
| $Q_o$ | Initial ionization in liquid produced by a 279 keV photoelectron. |
| $Q_o'$ | Initial ionization reduced by electron capture processes that do not depend on the strength of the electric field. |
| $Q_1(r)$ | Initial ionization escaping recombination. |
| $Q_2(r)$ | Initial ionization escaping recombination and capture by impurities. |
| $Q_3(r,r')$ | Amount of electron (and $Xe^+$ ion) charge produced in the avalanche as the electron charge $Q_2(r)$ drifts from r to r'. |
| $Q_4$ | Charge induced at charge amplifier by motion of initial ionization electrons (averaged over all photoelectron positions r). |
| $Q^-$ | Charge induced at charge amplifier by motion of avalanche electrons (averaged over all photoelectron positions). |
| $Q^+(T)$ | Charge induced at charge amplifier by motion of avalanche $Xe^+$ ions during time T (averaged over all photoelectron positions). |
| K | Recombination coefficient. |
| $\lambda$ | Coefficient of electron capture processes having a 1/E dependence. |
| $r_1$ | Radius of anode wire. |
| $r_2$ | Radius of cathode (4 mm for this experiment). |
| V | Anode-cathode potential. |
| E(r) | Strength of electric field. |
| T | Time during which charge motion can contribute to the pulse height at the output of the amplifier; determined by amplifier pulse shaping. |



When a 279-keV $\gamma$ ray produces a photoelectron in liquid Xe, the photoelectron is severely multiple Coulomb scattered in the liquid. It produces approximately 12 500 ion pairs distributed about the $\gamma$-ray interaction point within a distance of 240 $\mu$.[18] We shall therefore consider each photoelectron to produce a point-like packet of initial ionization at some distance r from the central axis of the chamber. Since the beam of 279-keV $\gamma$ rays was much narrower than the chamber (1.2 mm diam vs 8 mm diam) and attenuation was small, the distribution of values of r is fairly uniform. Consequently, after evaluating the expression for the pulse height as a function of photoelectron position r, initial ionization $Q_o$, recombination, electron attachment, and first Townsend coefficient we will average over all r.

The amount of electron charge $Q_1$ that escapes recombination may be approximated by

$$Q_1(r) = \frac{Q_o}{1 + K/E(r)}, \tag{4.1}$$

where $E(r) = V/[r\ln(r_2/r_1)]$, r is the radial coordinate of the $\gamma$-ray interaction point, $Q_o$ is the initial ionization, E(r) is the electric field at the position of the photoelectron, and K is the "recombination coefficient. This formula was used by Marshall[4] and is based on the Jaffe theory.[19]

As the electron charge $Q_1$ drifts toward the center of the chamber, it is subject to capture by electronegative impurities. The amount of free electron charge $Q_2$ reaching the central region of the chamber is given by

$$Q_2(r) = Q_1(r) \exp\left[\int_{r_1}^{r} -\sigma(r)dr\right], \tag{4.2}$$

where r is the radial position of the photoelectron, $r_1$ is the anode radius, and $\sigma$ is the probability of electron capture per cm of drift. Swan studied the dependence of $\sigma$ on electric field (E) for small amounts of oxygen in liquid argon and found that $\sigma = (8 \times 10^{9}\ V/cm^{2})\ F/E$, where F is the fraction of oxygen.[7] Since we are dealing with unknown contaminants in liquid xenon (rather than oxygen in liquid argon) we choose to approximate $\sigma(r)$ by the more general formula

$$\sigma(r) = \sigma_0 + \lambda/E(r), \text{ where } \sigma_0 \text{ and } \lambda \text{ are appropriate positive constants} \qquad (4.3)$$

We can find the integrated electron capture by using the $r^{-1}$ dependence of the electric field:

$$\int_{r_1}^{r} -\sigma(r)dr = -r\left[\sigma_o + \frac{\lambda}{2E(r)}\right], \qquad (4.4)$$

where we have neglected terms of order $r_1/r$. Substituting the evaluated integral into Eq. (4.2), we find

$$Q_2(r) = Q_o \exp(-r\sigma_o) \exp\left[-r\lambda/2E(r)\right] \Big/ \left[1 + K/E(r)\right]$$

$$= Q_o' \exp\left[-r\lambda/2E(r)\right] \Big/ \left[1 + K/E(r)\right], \qquad (4.5)$$

where we have absorbed into the coefficient $Q_o'$ those electron attachment terms that do not depend on the electric field.

It can be seen in Fig. 4 that below the onset of avalanche (i.e., below 1200 V) the three experimental curves differ by multiplicative factors (between 0.9 and 1.1) and that these factors do not vary by any significant amount (<2%) from 100 to 1200 V. Consequently, $Q_o'$ was allowed to take different values for each of the 2.9-, 3.5-, and 5.0-μ runs, while k and $\lambda$, parameters that control voltage-dependent factors,



were each required to take only a single value for all the data.

As the electron charge $Q_2$ approaches the fine central wire, it begins to avalanche. When the resulting electron swarm has reached distance r' from the center of the wire, its size is given by integrating Eq (1.1):

$$Q_3(r,r') = Q_2(r) \exp \int_{r}^{r'} \alpha(r'')\, dr''. \qquad (4.6)$$

The remaining analysis consists of calculating the charge induced at the anode by the motion of the ionization charge $Q_2(r)$ and the avalanche charge $Q_3(r, r')$.

In general, as a charge dq drops through a potential difference dV it induces a charge $d^2Q$ on the anode given by[20]

$$d^2Q = dq\,\frac{dV}{V}. \qquad (4.7)$$

For a charge moving from r to $r_f$ in a chamber of cylindrical geometry this becomes, after integration,

$$dQ = dq\,\frac{\ln(r/r_f)}{\ln(r_2/r_1)}. \qquad (4.8)$$

For small electron capture, the charge induced at the anode by the motion of the *initial ionization electrons* (averaging over all photoelectron positions) may be closely approximated by

$$Q_4 = \int_{r_1}^{r_2} \frac{Q_2(r)\ln(r/r_1)}{r_2\ln(r_2/r_1)}\, dr. \qquad (4.9)$$

The avalanche produces a pulse at the charge-sensitive amplifier due to the drift of the *avalanche electrons* toward the wire and by the drift of the *avalanche* $Xe^+$ *ions* away from the wire. The avalanche electrons drift at their saturated velocity [16] of $3 \times 10^5$ cm/sec and reach the surface of the anode in a few nanoseconds. On

the other hand the avalanche $Xe^+$ ions travel very slowly, requiring approximately 1 sec to drift from the anode to the cathode. Their mobility[21] is $3 \times 10^{-4}$ cm$^2$ volt$^{-1}$ sec.$^{-1}$ To have an acceptable signal-to-noise ratio, it is necessary to accept charge only during some small time interval T; i.e., to filter out the low-frequency noise components. Hence the filtered pulse from the amplifier is produced by the drift of the $e^-$ and $Xe^+$ ions only during the time T. Our amplifier and filter (1-μsec integration, 5-μsec differentiation) responded to a charge impulse (of 10-nsec width) with an output pulse that peaked 1.8 μsec after the initial rise. The electron avalanche pulses peaked in 2.0 μsec and were thus in effect only slightly slower than instantaneous pulses.

For $10^{-7}$ sec < T < $10^{-1}$ sec the filtered pulse contains all of the avalanche $e^-$ contribution but only a fraction of the $Xe^+$ contribution. It has been shown[22] that during a time T in a chamber of cylindrical geometry the $Xe^+$ ions produced in the avalanche will travel from r to $r^+$, given by

$$r^+ = \sqrt{r^2 + \frac{2\mu TV}{\ln(r_2/r_1)}}\ , \qquad (4.10)$$

where μ is the mobility of the $Xe^+$ ions, $r_1$ is the anode radius, and V is the potential on the chamber.

From Eq. (1.1), the additional electron (and $Xe^+$ ion) charge generated by the avalanche between r' and r' - dr' is given by

$$dQ_3(r, r') = Q_3(r, r')\alpha(r')dr'. \qquad (4.11)$$

To calculate the charge $Q^-$ induced at the anode by all avalanche electrons we multiply Eq. (4.11) (the differential electron charge) by Eq. (4.8) (the charge fraction induced when electrons drift from r' to the anode at $r_1$), integrate over r' (to include the entire avalanche), and finally average over all photoelectron positions r:



$$Q^- = \int_{r_1}^{r_2}\int_{r_1}^{r_2} \frac{\ln(r'/r_1)}{r_2\ln(r_2/r_1)} Q_3(r,r')\alpha(r')dr'dr. \qquad (4.12)$$

To calculate the charge $Q^+(T)$ induced at the anode by all avalanche $Xe^+$ ions, we multiply Eq. (4.11) (the differential $Xe^+$ ion charge) by Eq. (4.8) (the charge fraction induced when $Xe^+$ ions drift outward for time T), integrate over r', and average over r:

$$Q^+(T) = \int_{r_1}^{r_2}\int_{r_1}^{r_2} \frac{\ln(r'/r^+)}{r_2\ln(r_2/r_1)} Q_3(r,r')\alpha(r')dr'dr. \qquad (4.13)$$

## V. RESULTS IN LIQUID XENON

The sum of Eqs. (4.9), (4.12) and (4.13) were fitted to the data of Fig. 4 by adjusting the parameters $Q_o'$, K, λ and nine independent values of α for E = 400 kV/cm to 2000 kV/cm in steps of 200 kV/cm. The integrals were computed numerically. The best fit values are shown in Table II. Although $Q_o'$ was allowed to take a different value for each of the 2.9-, 3.5-, and 5.0-μ runs to allow for differences in the level of electronegative impurities and possible calibration errors, K, λ, and α(E) were required to fit the data for all three runs simultaneously. The fitting program used a linear interpolation on the table of log α vs E to define the function α(E) in Eq. (4.6). Since the pulses peaked 2 μsec after their rise, we fixed T at 2 μsec. When the fit was repeated assuming a value of 1.5 μsec for T, the best fit values of each parameter changed by less than 0.15 standard deviation. From measurements of the average ionization potential of gaseous Xe (21.9 eV/ion pair),[23] a 279-keV photoelectron should produce 2.04 X$10^{-15}$C of initial ionization ($Q_o$). We see from Table II that the average of the fitted values of $Q_o'$ is only 3% smaller than the expected value of $Q_o$, indicating that the loss of

electrons to impurities was a minor effect. Moreover, the best fit value for $\lambda$, the field-dependent attachment coefficient, was zero with a one-standard-deviation upper limit of 10 V/cm$^2$. Swan found[7] that for 1 ppm $O_2$ in liquid argon, $\lambda = 8 \times 10^3$ V/cm$^2$.

We assigned a ±10% uncertainty to each experimental datum point and the $\chi^2$ fit yielded the parameters shown in Table II. The data above 200 X10$^{-15}$C were omitted from the fit in order to exclude saturation effects. The resulting values of $\alpha(E)$ agree with those determined by the simple method described in Section III. The data for each rum are plotted in Fig. 5 with the curves from the fit, and the agreement is very good. The best fit $\chi^2$ was 84.0 with 111 degrees of freedom, verifying the value of the uncertainty (±10%) assigned to the data.

In Fig. 6 we plot the pulse induced on the charge amplifier as a function of photoelectron position r for several voltages in the 2.9$^-$μ run. These curves are the sum of the integrands in Eqs. (4.9), (4.12), and (4.13), before the averaging over r.

In Fig. 7 we plot the total charge G produced in the avalanche, the charge $G^-$ induced by the motion of the avalanche $e^-$, and the charge $G^+$ induced by the motion of the avalanche $Xe^+$ ions during the first 2 μsec as a function of voltage for the 2.9- and 5.0-μ wires. These three curves are based on the parameters shown in Table II and normalized to one initial electron.

For this calculation, we define:

$$G = \exp \int_{r_1}^{r_2} \alpha(r'')\, dr'', \tag{5.1}$$

$$G^- = \int_{r_1}^{r_2} \frac{\ln (r'/r_1)}{\ln (r_2/r_1)} \left[ \exp \int_{r'}^{r_2} \alpha(r'')dr'' \right] \alpha(r')dr', \tag{5.2}$$



$$G^+ = \int_{r_1}^{r_2} \frac{\ln(r^+/r')}{\ln(r_2/r_1)} \left[ \exp \int_{r'}^{r_2} \alpha(r'')dr'' \right] \alpha(r')dr' . \tag{5.3}$$

We see that of the charge produced in the avalanche (G) only a small fraction ($G^- + G^+$) is actually observed. This fraction varied from 0.10 to 0.19 during this experiment, increasing either with increasing voltage or with decreasing wire diameter. This behavior is quite reasonable, since an increase in the field at the anode causes the $Xe^+$ ions to travel a greater distance per microsecond. Fig. 7 also shows that the avalanche $e^-$ contribute much less than the avalanche $Xe^+$. As in gas-filled proportional chambers, this is due to the fact that the avalanche occurs close to the anode and the avalanche electrons fall through relatively few equipotentials before they reach the anode.

In Fig. 8 we plot the ionization coefficient $\eta = \alpha/E$ as a function of $E/\rho$ for liquid and gaseous xenon (where $\rho$ is the density in g/cm$^3$). The curve drawn through the liquid-data points is the gas-data curve displaced to the left by a factor of 26.

The only published measurement of $\alpha(E)$ in xenon gas was that of Kruithof.[24] His experiment consisted of measuring the average current in a parallel-plate chamber as a function of electric field. Kruithofs' values of $\eta = \alpha/E$ (shown in Fig. 8) all lie on a very smooth curve, even though the data were taken at many different pressures, ranging from 0.28 to 165 Torr. We regard this as a very accurate demonstration of $E/\rho$ scaling throughout the low-density region. We conclude that electrons in liquid Xenon do not avalanche as though they were in a gas having the density of liquid Xenon (3057 g/liter) but rather an effective density approximately 26 times smaller, assuming that $\eta$ is a function of $E/\rho$ only.

One might speculate that the avalanche was occurring in some gas-liquid mixture near the surface of the wire, to account for the lack of density scaling. The existence of such a mixture might be suggested by the strong pressure dependence seen in the avalanche of ionization pulses in liquid argon.[9] However, the pulse height in liquid xenon, unlike that in liquid argon, is insensitive to pressure changes of the order of 1 to 2 atmospheres. It is difficult to see how a gas-liquid mixture could exist at the wire even at 1 atm external pressure, since the presence of $10^6$ V/cm at the surface of the wire, along with the high gradient of the field, causes an additional pressure of approximately 1 atm on the liquid in the vicinity of the wire.

We find no evidence in the photopeak pulses for a second Townsend process. Photoelectrons liberated at the cathode by the prompt uv emission from excited Xe atoms would produce a second pulse delayed by 1.5 μsec, while electrons liberated at the cathode by the arrival of $Xe^+$ ions would produce a second pulse delayed by 1 sec. Such effects were not seen as multi-step pulses or as a background on the pulse-height analyzer. To investigate quantitatively the effect of possible second Townsend processes, we refit the data, assuming that for each ion pair generated in the avalanche, $3 \times 10^{-4}$ new electrons were *promptly* liberated in the chamber and that these electrons initiated a fresh avalanche. This number was chosen so that the largest pulse heights shown in Fig. 4 would have about a 30% contribution from the second Townsend processes. If the secondary electrons originate near the cathode, then this assumption is pessimistic, as the second and subsequent avalanches would be delayed, producing an easily detectable multi-step pulse shape in contradiction with our observations. After



this fit, the only parameter that was changed from its value in Table II by more than 0.25 standard deviation was the first Townsend coefficient for $2 \times 10^6$ V/cm, and this was decreased by only 12% (1.8 standard deviations).

At voltages higher than those used in this experiment (as described in Section III), secondary processes dominate by producing spurious pulses of irregular shape and size. Such pulses were not seen at the lower voltage settings used in the data-taking phase of the experiment. Even had they been present, they would not have shifted the photopeak but would have added to the background on the pulse-height analyzer.

## VI. RESULTS IN GASEOUS XENON

We have measured $\alpha(E)$ in pure gaseous xenon, using the 3.5- and 5.0-μ-diameter anode wires to check whether our technique (which uses detector pulses in a nonuniform field) would yield the same values as those of Kruithof[24] (who used direct current and a uniform field). Our measurement was performed by using the 5.9-keV x rays from an $^{55}Fe$ source and at a Xe gas pressure of 2.2 atm.[25] The resulting values of $\eta$ for gaseous Xe fall slightly below those of Kruithof (see Fig. 8).

## VII. DISCUSSION

In both liquid and gaseous Xe, the ionization coefficient $\eta$ rises rapidly as a function of E, then levels off. We may view $\eta$ as a direct measure of the efficiency for converting electron kinetic energy into ion pairs. If all the kinetic energy were used to create pairs with maximum efficiency, then $\eta$ would be 0.0824 ion pairs/eV, corresponding to the ionization potential of 12.13 eV.

When an energetic electron (such as a 279-keV photoelectron) passes through Xe it loses energy through mechanisms other than ionization. As a result, the average ionization potential is higher:

21.9 eV/ion pair, corresponding to an $\eta$ of 0.0457. We see from Fig. 8 that electrons heated to ionizing energies by an electric field are almost as efficient in creating ion pairs as are energetic electrons.

At present there is no theory of electron avalanche in dielectric liquids with which we can compare our experimental results. A few qualitative arguments, however, may serve to explain why electrons avalanche in liquid Xe more readily than in any other liquid studied to date. At very low electric fields, free electrons are in thermal equilibrium with the liquid and their drift velocity $v_d$ is proportional to the electric field E. According to the qualitative theory of Shockley[26] and of Staub,[27] an increase in electron agitation energy (average instantaneous kinetic energy) is accompanied by a decrease in the electron mobility $\mu$ (defined as $v_d/E$). This behavior is seen in Fig. 9 for the liquified noble gasses [16,28] but not for the liquid hydrocarbons.[29] Electrons in polyatomic liquids suffer inelastic collisions as they excite vibrational and rotational states and as evidenced by their high mobility are apparently prevented from acquiring any significant agitation energy, even at fields of $10^5$ V/cm. Of all the liquids for which we have measurements, liquid Xe has the highest ratio between its low-field and high-field mobilities. This high ratio is perhaps related to the ease of producing avalanche in liquid Xe.

Recently, Pisarev et al. have reported on a cylindrical proportional chamber filled with crystalline Xe.[30] Their chamber had a cathode diameter of 6 mm, an anode diameter of 10 $\mu$, and reached a gain of 150 at 5200 V. Using the values of $\alpha(E)$ measured in this experiment for liquid Xe and the positive ion (hole) mobility of 0.02 $cm^2\ V^{-1}\ sec^{-1}$ measured for



solid Xe,[16] we would predict a gain of 150 at 4100 V, in good agreement with the measured voltage. We conclude that the avalanche in solid Xe is very similar to that in liquid Xe.

Pisarev et. al., have also measured pulse height vs. voltage in the same chamber filled with gaseous xenon at pressures ranging from 1 to 50 atm.[31] Their solid-xenon data agree best with their gaseous-xenon data for 30 atm, showing that the required electric field for electron avalanche is similar in both gaseous xenon at a density of $\sim$240 g/liter and solid xenon at the much higher density of 3540 g/liter.

## VIII. CONCLUSIONS

We find that the avalanche of free electrons in purified liquid Xe is adequately described by the same formalism that is used for the avalanche of electrons in gases. Although the ionization coefficient $\eta\ (E/\rho)$ in liquid Xe has a shape similar to that measured in gaseous Xe, it is not the same function in the sense that strict "density scaling" does not apply. In gasses, $\eta$ is a function of $E/\rho$ over a large range of densities, but the function $\eta\ (E/\rho)$ measured in gaseous Xe is valid in liquid Xe only when the electric field is scaled by an "effective density" of 117 g/liter rather than the true density of 3057 g/liter.

## ACKNOWLEDGMENTS

We are indebted to Joe Savignano, Tony Vuletich, and Buck Buckingham for their skill and patience in building and maintaining our equipment. We are grateful to Robert Flagg, Steven Louie, and Robbie Smits for assistance in data taking and to Leo Fallicov and Sherwood Parker for many helpful discussions. We thank Dane Anderberg for his excellent glass work and the Inorganic Materials Research Division of LBL for the use of their scanning electron microscope.

FOOTNOTES AND REFERENCES

* Work supported by the U.S. Atomic Energy Commission and the National Aeronautics and Space Administration.

1. N. Davidson and A. E. Larsh, Jr., Phys. Rev. 74, 220, (1948); 77, 706 (1950).

2. G. W. Hutchinson, Nature 162, 610 (1948).

3. M. S. Malkin and H. L. Schultz, Phys. Rev. 83, 1051 (1951).

4. John Marshall, Rev. Sci. Instr. 25, 232 (1954).

5. Due to the formation of a "bubble" around the electron in liquid He [L. Meyer and F. Reif, Phys. Rev. 110, 279 (1958) and R. L. Williams, Can. J. Phys. 35, 134 (1957),] and possibly in liquid Ne [L. Bruschi, G. Mazzi, and M. Santini, Phys. Rev. Letters 28, 1504 (1972)] the electron mobility in these liquids is considerably lower than in the liquids argon, krypton, and xenon.

6. D. W. Swan, Proc. Phys. Soc. 76, 36 (1960).

7. D. W. Swan, Proc. Phys. Soc. 83, 659 (1964).

8. L. W. Alvarez, Lawrence Radiation Laboratory Group A Physics Note 672, Nov. 26, 1968 (unpublished).

9. S. E. Derenzo, D. B. Smith, R. G. Smits, H. Zaklad, L. W. Alvarez, and R. A. Muller, Lawrence Radiation Laboratory Report UCRL-20118, National Accelerator Laboratory Summer Study Report SS-181, July 1970, Batavia, Illinois.

10. R. A. Muller, S. E. Derenzo, G. Smadja, D. B. Smith, R. G. Smits, H. Zaklad, and L. W. Alvarez, Phys. Rev. Letters 27, 532 (1971).

11. S. E. Derenzo, R. Flagg, S. G. Louie, F. G. Mariam, T. S. Mast, R. A. Muller, A. J. Schwemin, R. G. Smits, H. Zaklad, and L. W. Alvarez, Lawrence Berkeley Laboratory Report LBL-1321, in *Proceedings of the XVI International Conference on High Energy Physics, Chicago and Batavia, Illinois, September 1972*, Vol. 2, p. 388.



12. In liquid argon, the electron avalanche can occur only at places scattered sparsely along the wire (see Ref. 9). One consequence is that only a small fraction of incident particles initiate an avalanche, an effect also observed by E. A. Kushnirenko and A. G. Chilingarov (*Proceedings of the International Conference on High Energy Physics Instrumentation, Dubna, U.S.S.R.*, Vol. 1, p. 297, December 1971). In liquid xenon the avalanche occurs with ∿100% probability along the length of the wire.

13. Our fine tungsten wire was purchased from the following suppliers:
    a) 2.9 μ: Sigmund Cohn Corp. 121 South Columbus Avenue, Mount Vernon, N.Y. 10553.
    b) 3.5 μ: Philips Elmet Corp. Lewiston, Maine 04240 (no longer can supply wire below 5 μ diam).
    c) 5.0 μ: Sylvania Metals and Chemicals Towanda, Penn. 18848.

14. Supplied by Buckbee-Mears, 245 East 6th Street, St. Paul, Minn. 55101.

15. H. Zaklad, Lawrence Radiation Laboratory Report UCRL-20690, April 1971 (unpublished).

16. L. S. Miller, S. Howe, and W. E. Spear, Phys. Rev. 166, 871 (1968).

17. Yu. V. Kulikov, N. I. Molashkevich, and V. D. Ruabtsov, JINR P13-5403, Dubna (1970).

18. The effective range of electrons in liquid xenon was assumed to be similar to that measured in Aluminum by J. S. Marshall and A. G. Ward, Can J. Res. A15, 39 (1937).

19. G. Jaffe, Le Radium 10, 136 (1913). For a simpler discussion, see A. N. Gerritsen, Physica 14, 381 (1948).

20. Hans H. Staub, *Experimental Nuclear Physics*, Vol I, edited by E. Segrè (John Wiley, New York, 1952), p. 16.

21. H. Davis, S. Rice, and L. Meyer, J. Chem. Phys. 37, 947 (1962).

22. H. H. Staub, p. 35 of Ref. 20.

23. I. T. Myers, *Radiation Dosimetry*, Vol. 1, Ch. 7, Ionization, 2nd ed. (Academic Press, New York 1968).

24. A. A. Kruithof, Physica 7, 519 (1940).

25. S. G. S. Louie, Lawrence Berkeley Laboratory Group A Physics Note 765, 1972 (unpublished).

26. W. Shockley, Bell System Technical Journal 30, 990 (1951).

27. H. H. Staub, p. 6 of Ref. 20.

28. D. W. Swan, Proc. Phys. Soc. 83, 659 (1964).

29. W. F. Schmidt and A. O. Allen, J. Chem. Phys. 52, 4788 (1970).

30. A. F. Pisarev, V. F. Pisarev, and G. S. Revenko, Dubna Report JINR-P13-6449 (to be published in JETP).

31. A. F. Pisarev, V. F. Pisarev and G. S. Revenko, Dubna Report JINR-P13-6450 (to be published in JETP).



TABLE II.

Parameters resulting from the fit described in the text.

| Quantity | Best Fit Value |
|---|---|
| $Q_o'$ (2.9-μ run) | $(2.22\pm0.14)\times10^{-15}$C |
| $Q_o'$ (3.5-μ run) | $(1.86\pm0.10)\times10^{-15}$C |
| $Q_o'$ (5.0-μ run) | $(1.96\pm0.14)\times10^{-15}$C |
| K | $98\ ^{+12}_{-10}$ V/cm |
| λ | $0\ ^{+10}_{-0}$ V/cm$^2$ |
| α (400 kV/cm) | $470\ ^{+600}_{-470}$ cm$^{-1}$ |
| α (600 kV/cm) | $8050\pm3500$ cm$^{-1}$ |
| α (800 kV/cm) | $2550\ ^{+7000}_{-2550}$ cm$^{-1}$ |
| α (1000 kV/cm) | $8300\ ^{+3400}_{-4700}$ cm$^{-1}$ |
| α (1200 kV/cm) | $15000\pm6000$ cm$^{-1}$ |
| α (1400 kV/cm) | $27500\pm6000$ cm$^{-1}$ |
| α (1600 kV/cm) | $33500\pm4500$ cm$^{-1}$ |
| α (1800 kV/cm) | $38400\pm4500$ cm$^{-1}$ |
| α (2000 kV/cm) | $44700\pm2600$ cm$^{-1}$ |

FIGURE CAPTIONS

Fig. 1. Liquid xenon single-wire proportional chamber. Cathode is 8 mm diameter, formed by depositing a conductive coating of SnO onto the inside glass wall. Central anode wires of 2.9-, 3.5-, and 5.0-μ-diameter tungsten were used. Chamber was connected directly to a charge amplifier.

FIG. 2. Scanning electron micrographs of the tungsten anode wires used in this experiment. The 2000 mesh per inch Buckbee-Meers screen (used for calibration) may be seen beneath the 5-μ wire.

FIG. 3. Pulse-height spectrum from a collimated $^{203}Hg$ source in chamber of Fig. 1 filled with liquid xenon, using a 3.5-μ anode at 2400 V, showing 68-keV photopeak, 279-keV Compton edge, and 279-keV photopeak. Gain in the liquid was 10 and the field at the anode surface was $1.8 \times 10^6$ V/cm.

FIG. 4. Pulse height of 279-keV photopeak vs voltage for 2.9-, 3.5-, and 5.0-μ-diameter anode wires.

FIG. 5. Data of Fig. 4 shown with best fit curves. See test for details.

FIG. 6. Calculated pulse size for a 2.9-μ anode as a function of photoelectron position for a series of applied voltages. Calculation based on best fit parameters shown in Table II.

FIG. 7. Calculated curves comparing the size of the avalanche with the size of the $e^-$ and $Xe^+$ ion contributions to the observed pulse. All curves are normalized to one initial $e^-$ and based on the best fit $\alpha(E)$ function shown in Table II. [See Eqs. (5.1) to (5.3) and associated text for further details.] Solid lines: charge of $e^-$ (and $Xe^+$) produced in the avalanche. Dot-dashed lines: charge induced by the avalanche $Xe^+$ ions in



the first 2 μsec. Dotted lines: charge induced by the avalanche electrons.

FIG. 8. Ionization coefficient vs the ratio of electric field to density for gaseous xenon (squares, Ref. 24; circles, this experiment) and for liquid xenon (circles, this experiment). The curve drawn through the liquid-data points is the gas data curve of Ref. 24 displaced to the left by a factor of 26.

FIG. 9. Electron drift velocities observed for various liquids as a function of electric field. Liquified noble gases at their normal boiling points (Refs. 5 and 16). Liquid hydrocarbons at 23°C (Ref. 29). Solid lines: observed data; dashed lines: linear extrapolations.

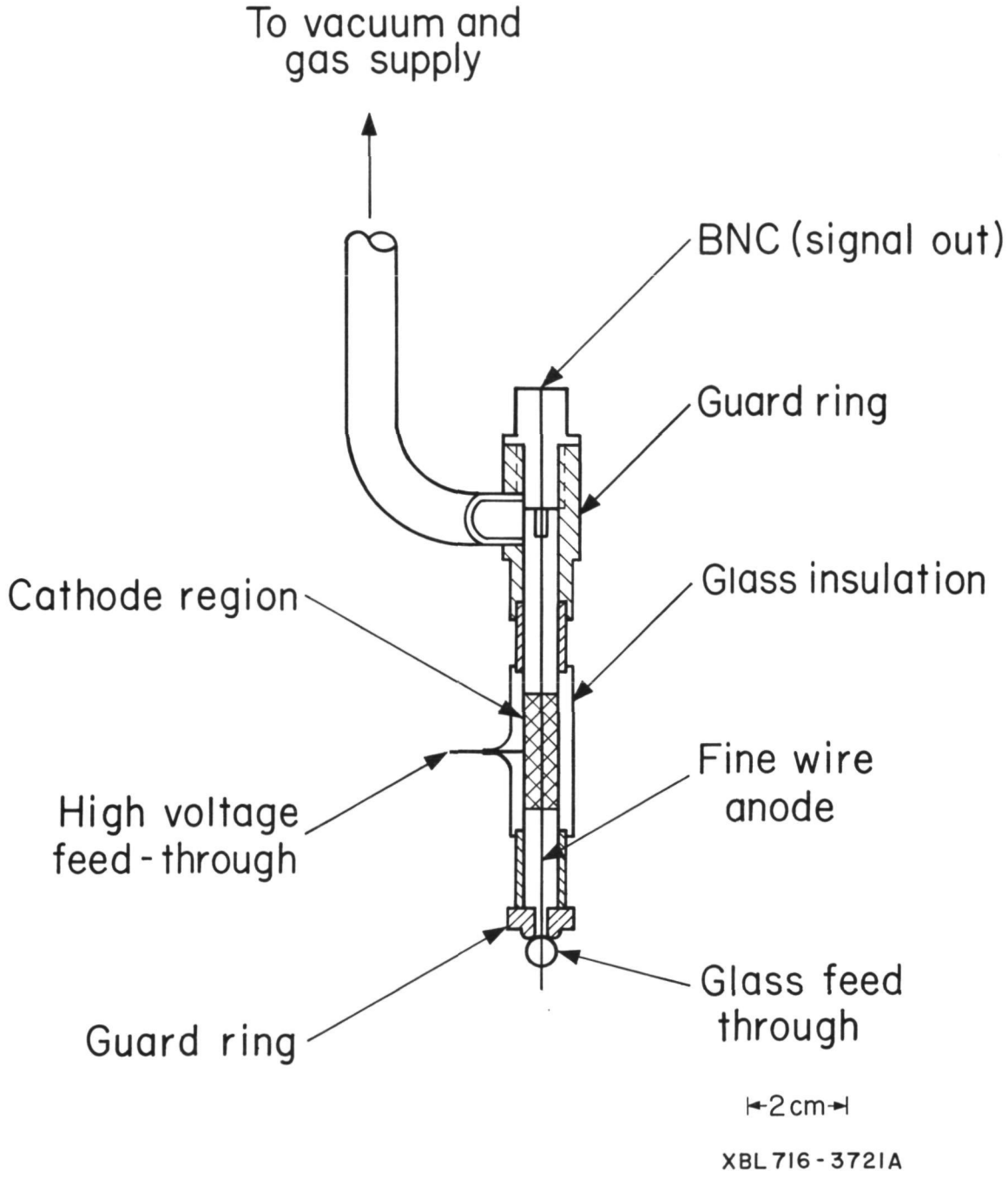
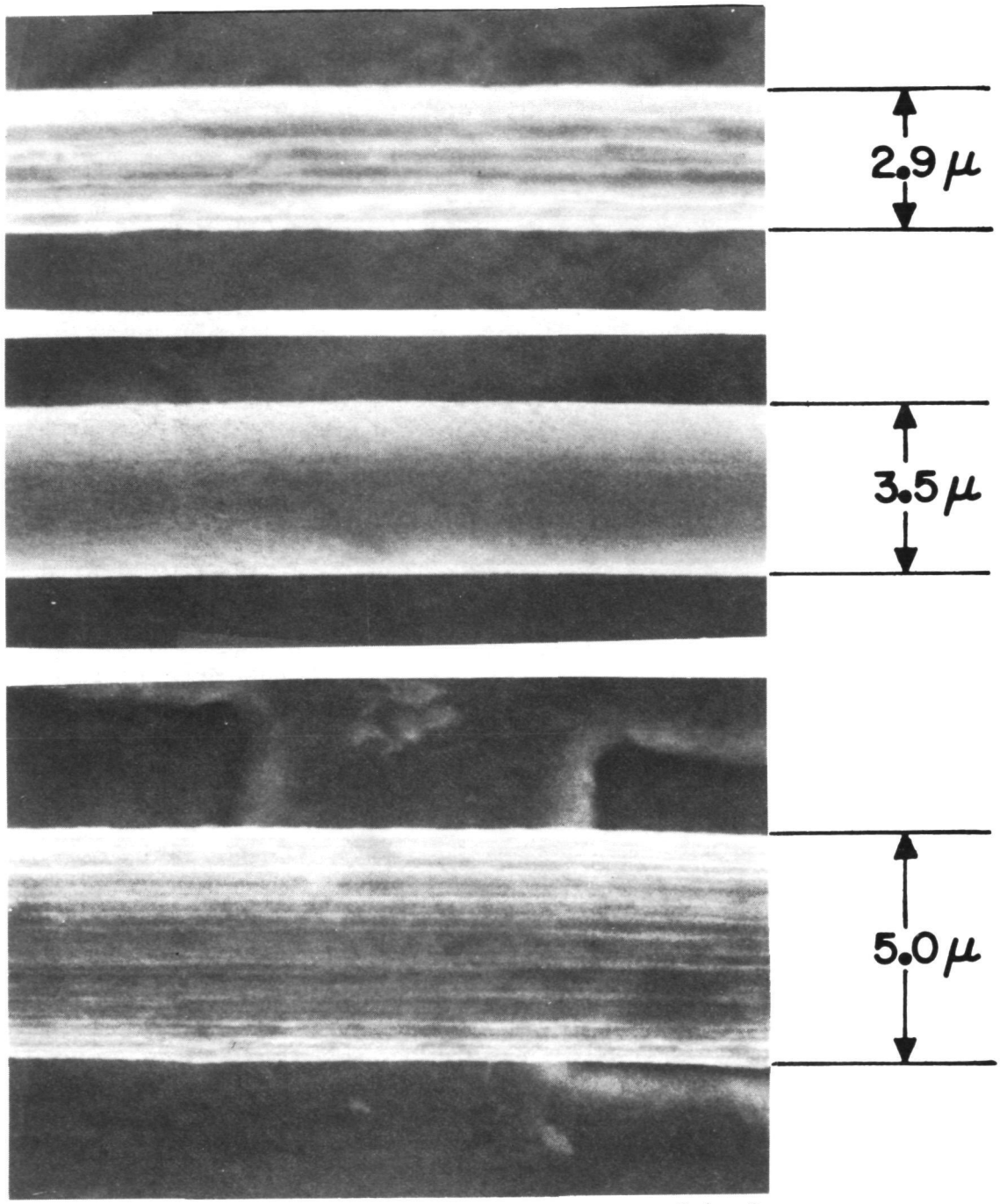

To vacuum and
gas supply
BNC (signal out)
Guard ring
Glass insulation
Cathode region
Fine wire
anode
High voltage
feed-through
Glass feed
through
Guard ring
2 cm
XBL 716-3721A

Fig. 1




XBB 728-4278

Fig. 2

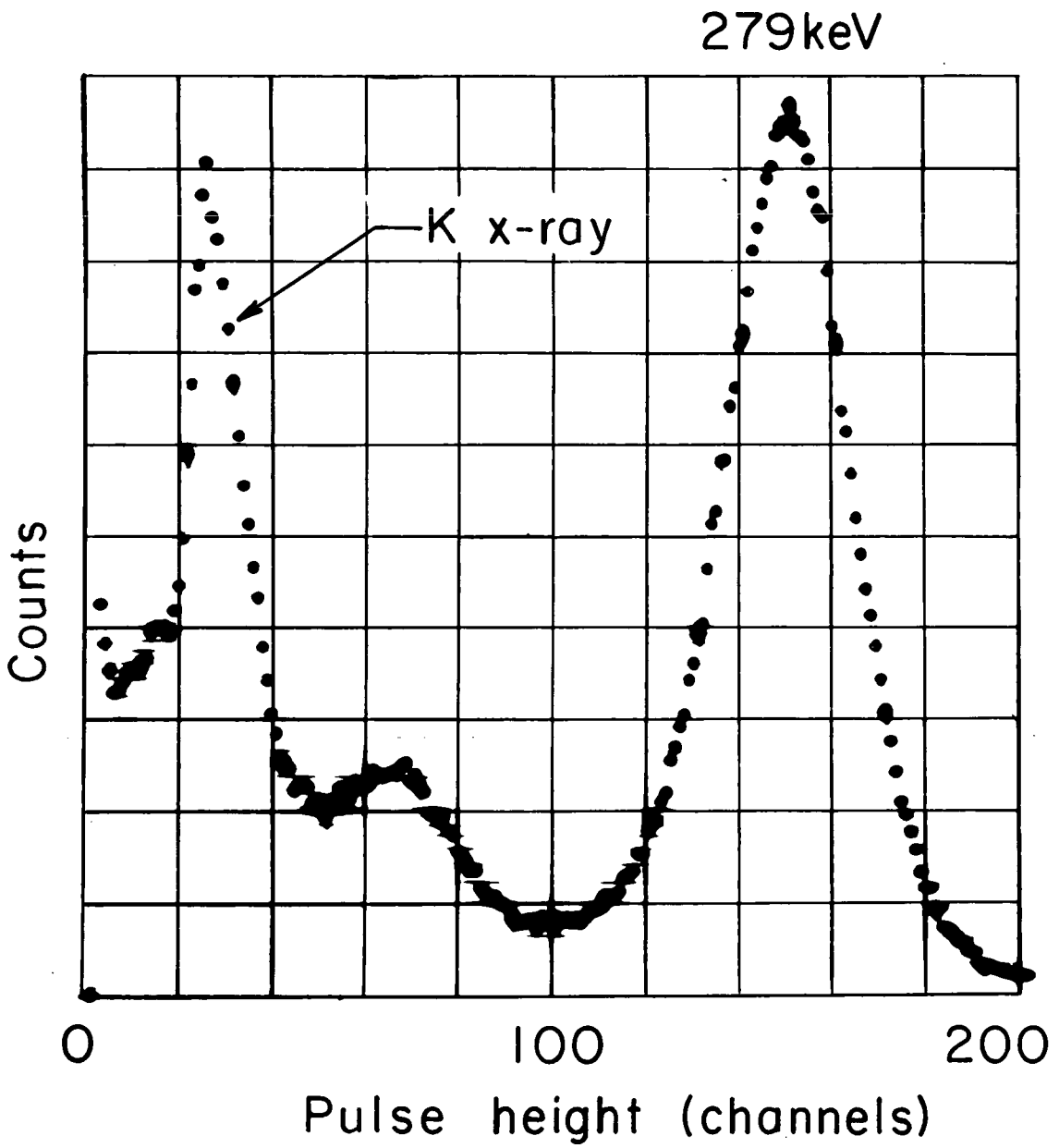
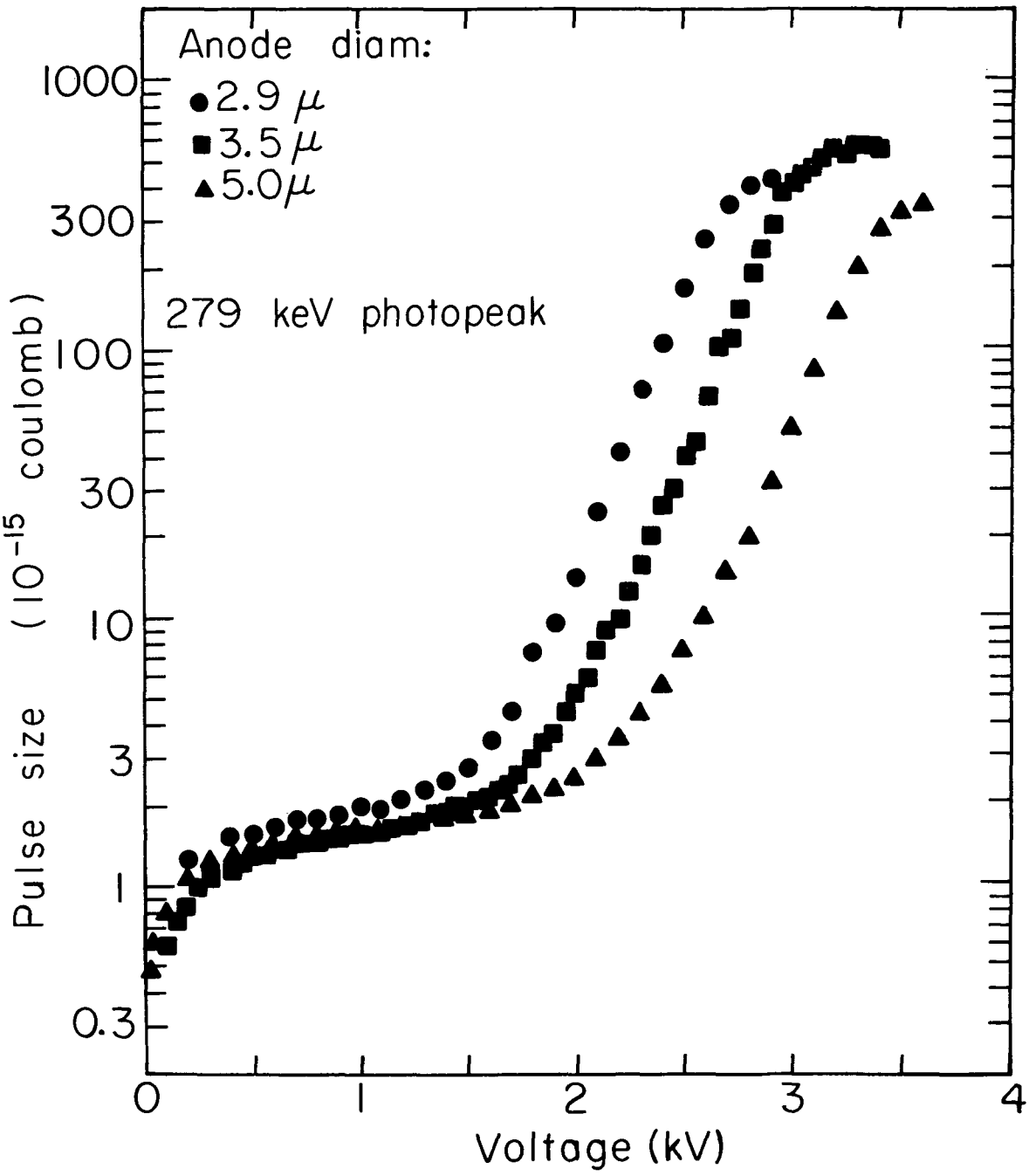

279keV
K x-ray
Counts
0
100
200
Pulse height (channels)
XBL722-2248

Fig. 3




XBL734-2639

Fig. 4

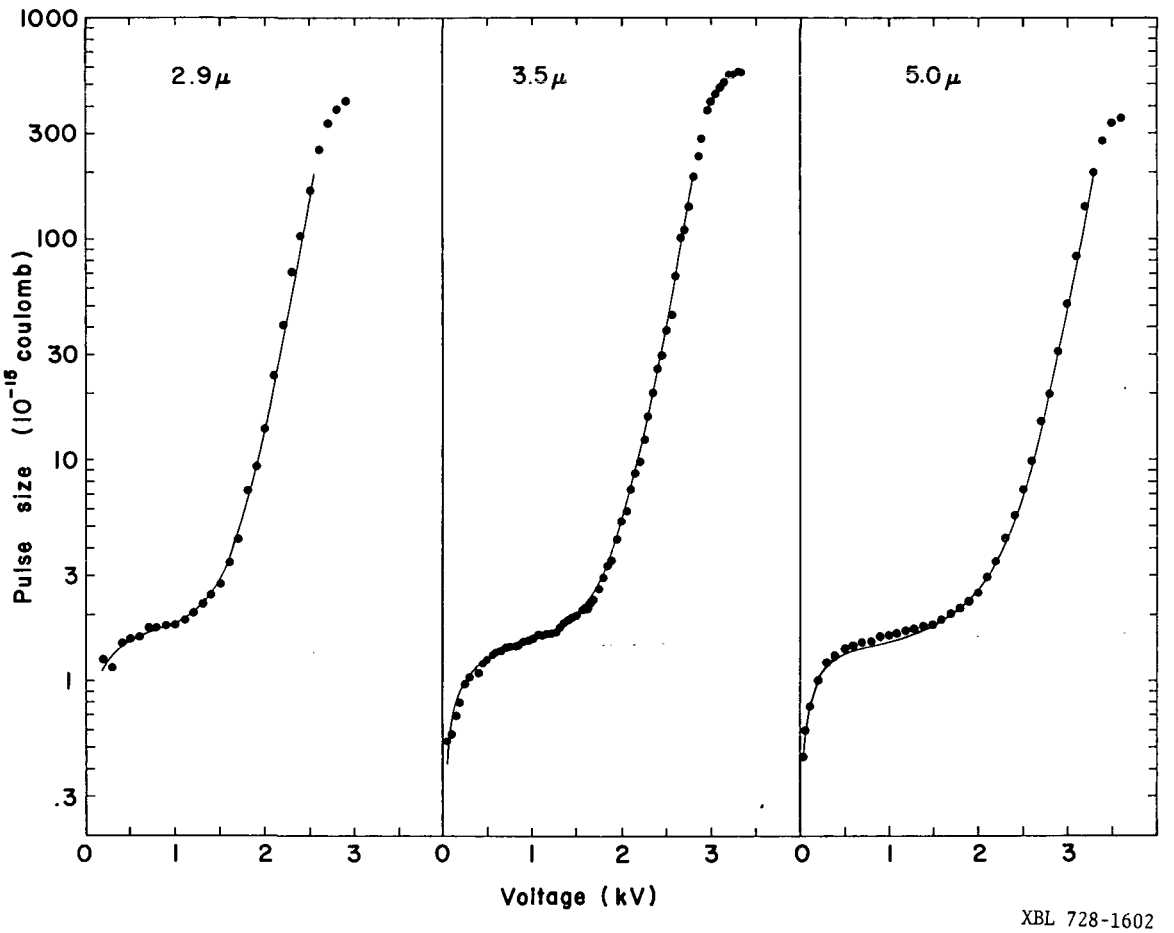
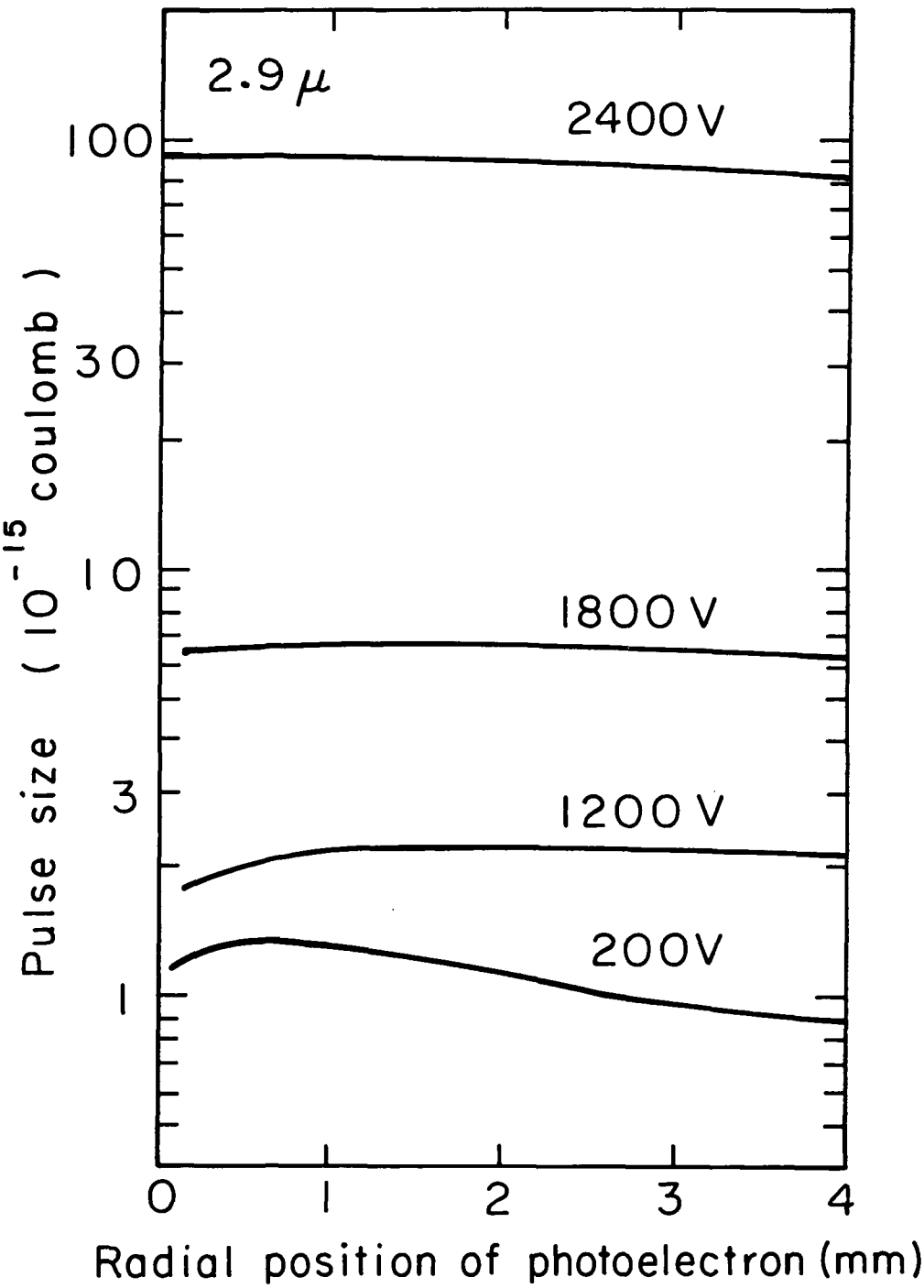
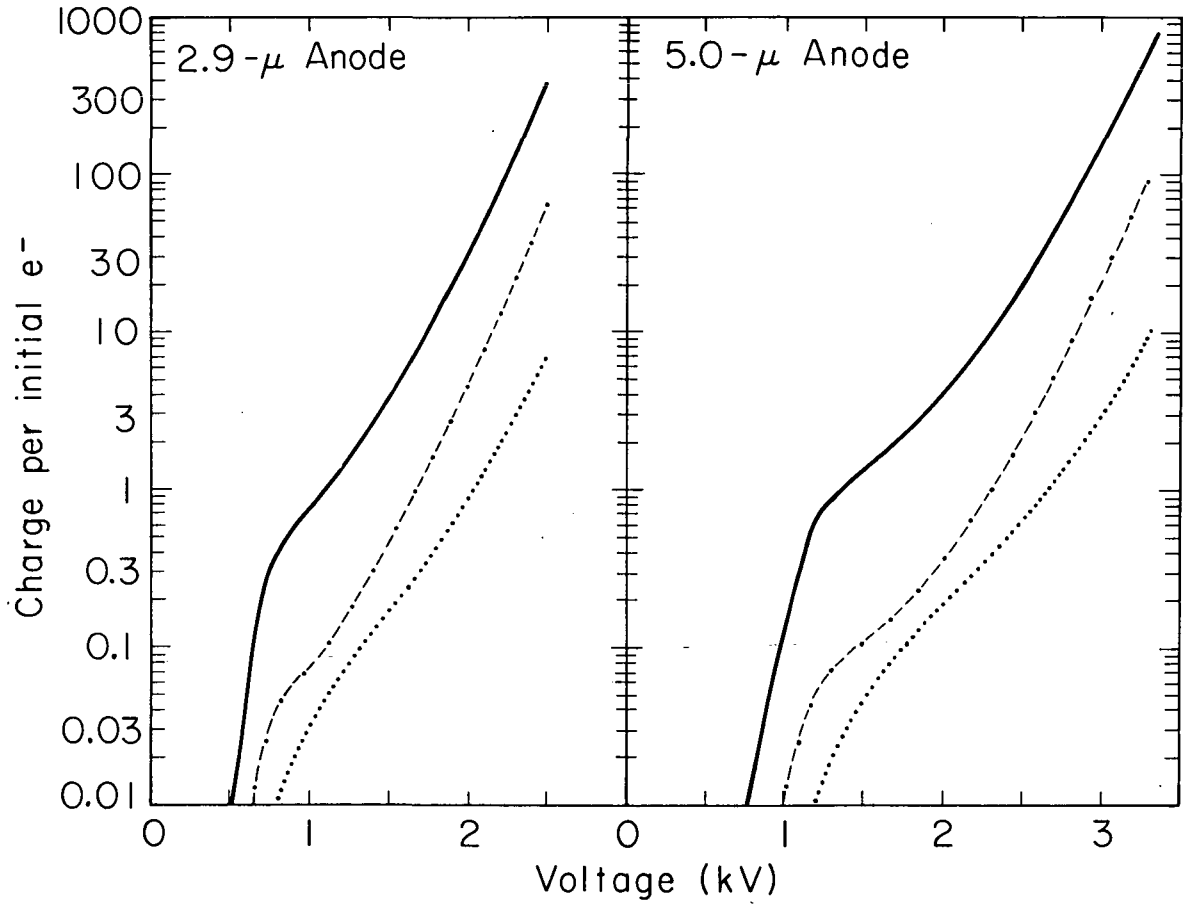

2.9μ
3.5μ
5.0μ
Pulse size ($10^{-15}$ coulomb)
1000
300
100
30
10
3
1
.3
0
1
2
3
Voltage (kV)
XBL 728-1602

Fig. 5



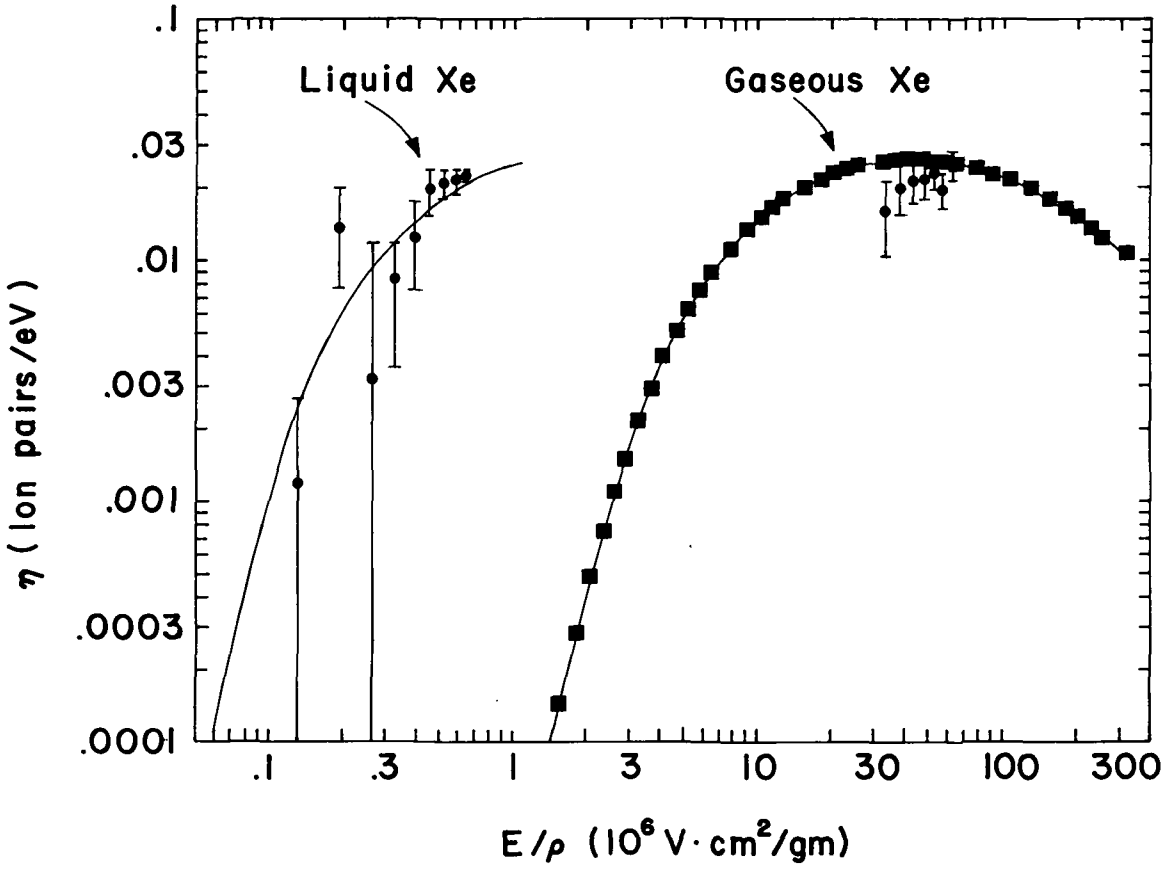

XBL734-2621

Fig. 6



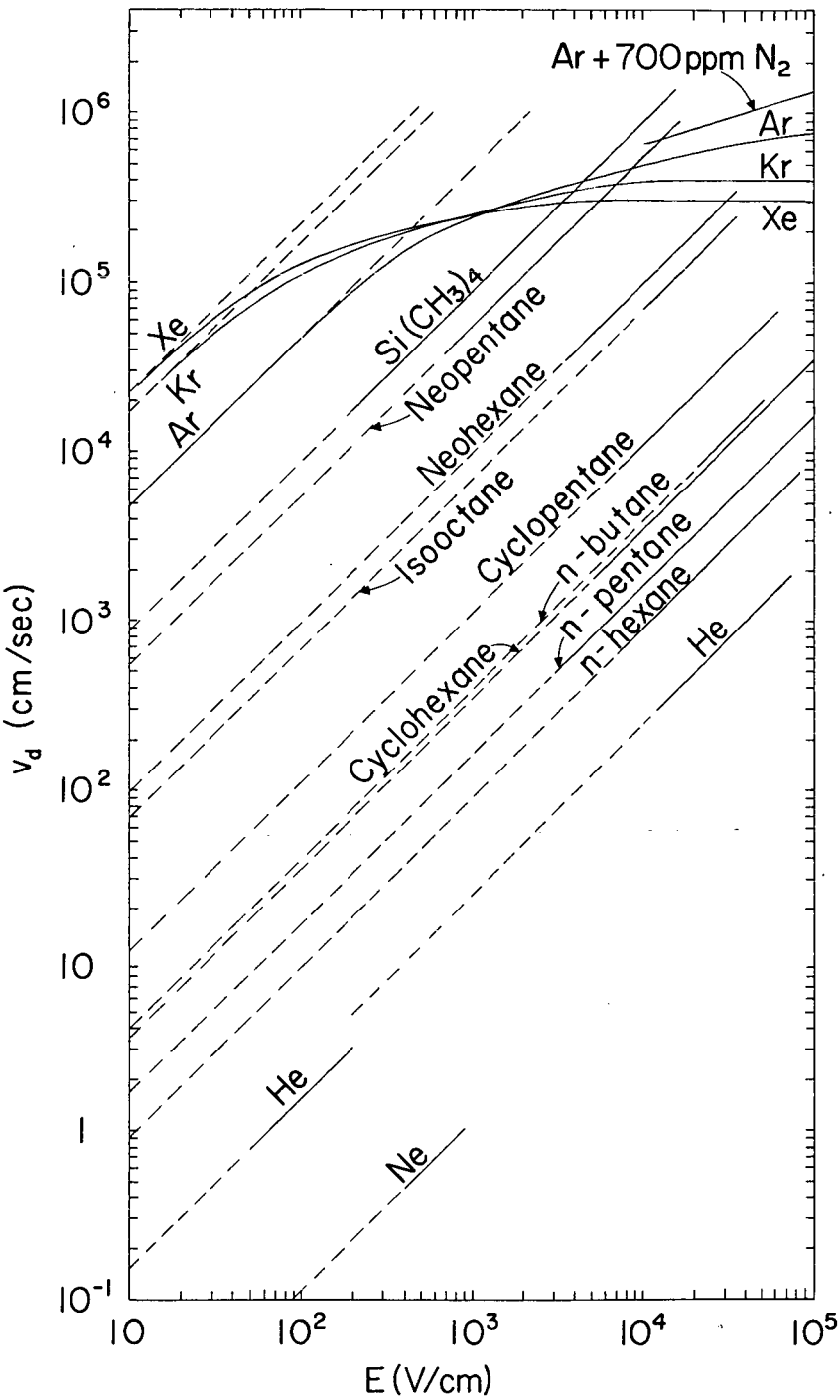

XBL734-2620

Fig. 7




XBL 728-1601

Fig. 8




XBL 719-4252A

Fig. 9

are added and the buffer contents are dumped onto one of the magnetic tape units. Tapes generated in this way contain all the scientific telemetry data plus time, position, date, and altitude data which are generated or calibrated on the ground. Each tape record is thus complete within itself and contains all the information necessary for analysis. The use of two tape drives with automatic crossover assures that no data are lost during tape-changing operations.

Other routines that are serviced at lower priorities include two data printout programs. One is used to calculate and print information that is important for balloon control and tracking purposes, including altitude in feet, rate of ascent in feet per minute, position in longitude and latitude, and time. The other provides a real-time look at scientific data in engineering units. The interval between printouts of these two routines is established by front-panel selector switches and can be changed at any time during a flight. In addition, a printout of either program can be requested at any time by the operator.

A down-range receiving station is used when the flight trajectory extends beyond the range of coverage from Palestine. This station consists of essentially the same equipment as at Palestine, with the exception of the computer. A manual version of the PCM decommutation equipment is available for real-time monitoring and conditioning of signals for analog recording. The analog tapes are played through the computerized station after the flight to generate digital tapes. This procedure gives the experimenter a complete set of industry-compatible tapes covering his data for the entire flight. These tapes can easily be read into a larger computer for data reduction.

## Flight Preparation

Considerable coordination is needed to make full use of the data handling capability outlined above. Each scientist's flight request includes his requirements for data retrieval. These requirements are reviewed by NSBF personnel and discussed with the scientist. A format is selected for the data encoder that will provide the necessary sampling rates for all parameters and the EPROM is ordered if necessary. Real-time monitoring and programming requirements are outlined several weeks before the scheduled flight date so that all software can be generated. The airborne hardware and decommutation patchboards are prepared for use when the experimenter arrives in Palestine. The data system then plays a very important role in the preflight checkout by monitoring all the parameters of the system. •

Atmospheric Technology 5, 25 (March 1974)

A72-27296

# Searching for Antimatter in the Cosmic Rays

*Andrew Buffington, University of California at Berkeley*

It is a great adventure to build an instrument to measure cosmic rays, to take it far from home to fly it by balloon, to return with data to analyze, and then to interpret the results. Cosmic ray studies are particularly fascinating because they use measurements of the smallest things in nature——the elementary particles——to deduce facts about the largest things in nature——the galaxy and the universe. Existence of antimatter is one of the most fundamental and striking symmetries discovered in these elementary particles. Our group has conducted a search to see whether antimatter particles can be found in the primary cosmic rays, and several times we have had the pleasure of using the balloon launch facilities at Palestine. I believe our search for antimatter is an exciting example of the motivation which brings scientists to the balloon facility, of the excitement and hard work they experience while preparing and flying their instruments, and of the knowledge that results from their measurements. Readers who would enjoy learning more about antimatter or about our group's other cosmic ray experiments should consult the sources listed at the end of this article in the section titled "For Further Reading."

## Antimatter in the Universe

Antimatter is a good example of a discovery from one branch of physics which stimulates profound thought in another branch of physics. Elementary-particle physicists found that when matter is produced by high energy collisions in accelerator experiments, an

equal amount of antimatter is always created at the same time. Soon after this symmetry was discovered, some astrophysicists began to wonder whether large amounts of antimatter were created billions of years ago in the early universe. If the rules of creation of matter were the same then as they are now in our laboratories, we should expect the universe to contain equal amounts of matter and antimatter. But where could this antimatter be? We believe it is absent here on earth and in the solar system, since we never see the tremendous annihilating explosions and accompanying gamma rays that would result from chance encounters of pieces of antimatter with matter. The low observed fluxes of gamma rays imply that matter and antimatter intermix very little throughout our galaxy—less than one part per billion.

Perhaps some sort of barrier exists which keeps matter and antimatter apart. Our galaxy could be composed half of matter and half of antimatter, with a complex array of these barriers preventing the mixing. We cannot tell whether this is the case by looking through telescopes, since the light from antimatter stars would be exactly the same as that from stars of "normal" matter. We would like to travel around the galaxy to gather samples of material for analysis, but that ideal seems to lie beyond technical reach.

Fortunately, material from far outside the solar system comes to us in the form of energetic cosmic rays. These particles, which were ejected millions of years ago by stars in violent late stages of their evolution, come to us primarily from the nearest 1% of the galaxy, although some have come from much farther away. Analyzing the cosmic rays could tell us whether there are regions of antimatter somewhere near us in the galaxy. If there are, we might someday be able to bring enough back to satisfy the apparently insatiable energy needs of earth. Such distant space travel seems to be very difficult, but given sufficient motivation perhaps a means could be found.

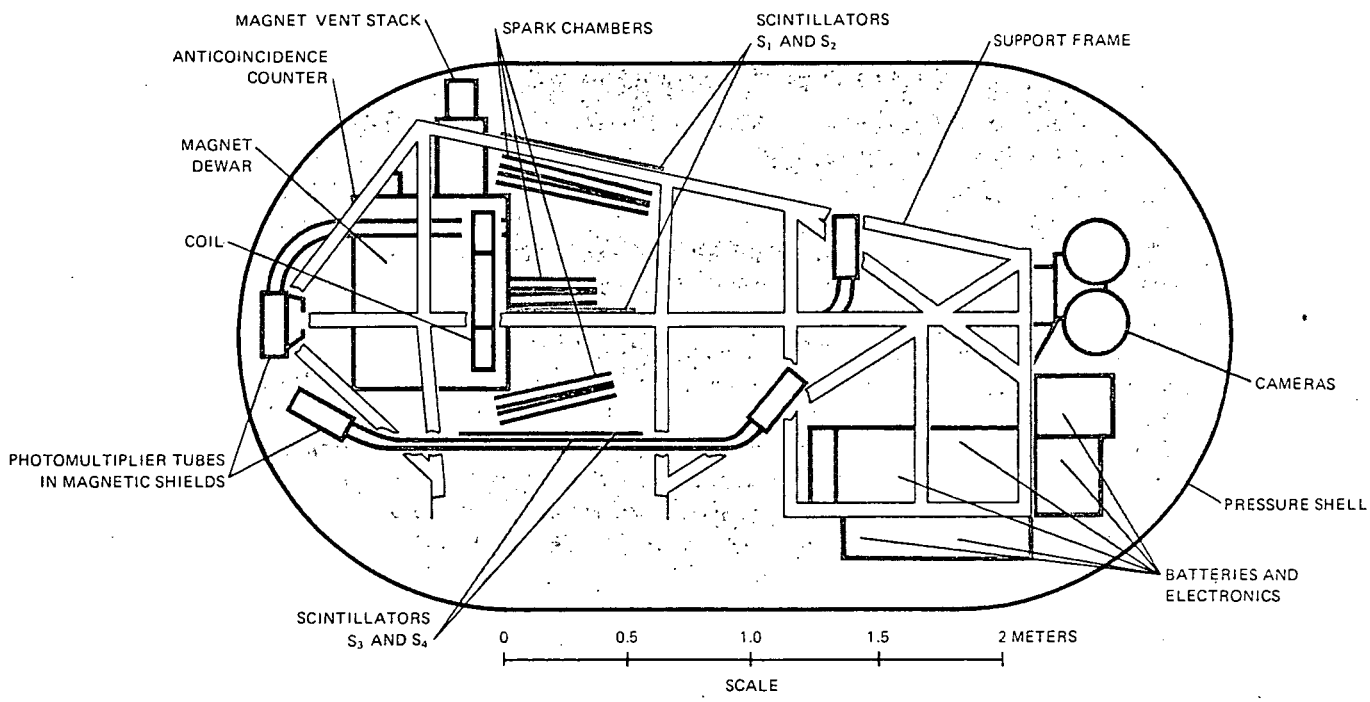
## Techniques

Antimatter has two properties that distinguish it from normal matter. The first of these, the well-known annihilation which results when antimatter touches matter, might someday make antimatter a good source of energy. This property was also used as a characteristic signature in the early searches for antimatter in cosmic rays: an emulsion block was flown above the atmosphere, and a large number of fairly low energy cosmic ray nuclei came to rest in it. The track of each stopping nucleus was examined for signs of violent annihilation explosions from an antinucleus. Various groups performed such searches several years ago; they saw no annihilation explosions in about 20,000 stopping cosmic rays.

However, the low energy cosmic rays examined by this technique would probably not penetrate a matter/antimatter barrier. We wanted to carry out our search at higher energies in hopes of getting a more representative sample of material from the galaxy. To accomplish this we used the other crucial property of antimatter: the negative electric charge of antinuclei is opposite to that of normal matter. This means that normal nuclei and antinuclei bend in opposite directions when they travel through a magnetic field.

To search for antimatter, we built a superconducting magnetic spectrometer, shown schematically in Fig. 1. The magnet, with its liquid helium cooling system, provides a strong magnetic field over a large volume. When an energetic cosmic ray comes from above, it passes through trigger scintillators. When the electronics viewing these scintillators

*Fig. 1 The superconducting magnetic spectrometer (courtesy of the Lawrence Berkeley Laboratory, University of Calif. [illegible].*



detects a highly charged nucleus that has passed all the way through the apparatus, it triggers the spark chambers. These mark the trajectory of the particle with bright sparks, whose locations are recorded on film by the camera. In a complete flight, we usually recorded 30,000 cosmic ray events with this technique. After the flight, we looked for oppositely bent trajectories that indicated negatively charged nuclei. The technique enabled us to look for antinuclei in cosmic rays with up to 50 times the energy allowed by the emulsion block technqiue.

Since antimatter cosmic rays were quickly destroyed by interactions in the atmosphere, we had to fly our instrument by balloon to be able to gather our data. The project was our first experiment carried out by remote control; all of our previous experience had been at accelerators where we had access to the experiment while it was in progress. It took our group about a year and a half to put together our complicated instrument with its cryogenic magnet, spark chambers, optics, electronics, scintillators, and pressure shell. Although the physicists and support staff each had various responsibilities in the construction and testing of the instrument, each of us, at some point, had to become intimately familiar with the whole system. To get his equipment working right, and to understand the measurements he makes with it, every experimental physicist has a "love affair" with his instrument. He becomes involved with it to the exclusion of almost everything else. For each flight of our instrument, this period of intense involvement extended over the final months of test and calibration at Berkeley, and culminated in our month's stay at Palestine.

## Flights from Palestine

A field operation with our gondola means many things to each member of our group. To everyone it means many days with long hours of work to bring the instrument to a state of proven flight readiness. To each member of the team it means time-consuming care in performing his repairs and calibrations, since the success of the flight will hinge on efficient performance of the components. Since our gondola is complex, we cannot afford to fly it often and must depend on a high percentage of success in our flights.

In Palestine we frequently see familiar things juxtaposed with new and unfamiliar things. Strange Texas creatures fly and climb into our optics and electronics boxes. Scorpions come out of hiding to menace us when we have the lights out for our optical calibrations. Never have I seen a place with bigger bugs than Texas! The location brings out new aspects in the members of our team; I hear our people telling new stories at lunch or to the NCAR crew in the coffee room. Also, every physicist feels the need to become an amateur meteorologist during his stay, and we all spend lots of time perusing the maps and crowding the weather briefings. The long days of preparation time test our nerves and gradually lower our tolerance to the absence of our families and urban comforts.

Flight day is the high point of our adventure. First we make the crucial checkout; then we attend the weather briefing, watch the balloon layout and inflation, and finally see our equipment launched. In our trailer excitement mounts during the ascent and early data-taking; each of us wonders whether all the arduous preparations have been successful or whether some failure will cheat us of our data and perhaps even of our instrument. We have feedback through telemetry, and each member of our team has to scan a portion of the information for anomalies. Through our extensive electronics redundancy and our command capability, we can "fly" our experiment, getting physics and engineering information in real time and altering our flight profile to meet unexpected trigger rates, electronics failures, and changing weather conditions. Our group has had very good luck with ballooning; we have obtained good physics data from five out of six flights.

After the flight several members of our group usually have enough energy left to go along with the recovery crew to find where the gondola has come down. To me, this is the strangest experience of all in ballooning, since the gondola looks very much out of place, whether it has landed in the woods, on the prairie, or in a farmer's field. Frequently a crowd seems to spring out of the ground and wants to know all about it. It generally takes a day or two to get the gondola loaded onto a truck and back to Palestine. Another week's work is required to "debrief" the equipment and to make it ready for return to Berkeley.

## Data Analysis and Results

After developing the film we measure the sparks on an automatic machine and

## For Further Reading


Alfven, H., 1966: *Worlds-Antiworlds: Antimatter in Cosmology*. W.H. Freeman, San Francisco.

Buffington, A., R. A. Muller, L. H. Smith, and G. F. Smoot, 1972: High energy particle astronomy. *Sci. Technol. 28*, 289 - 298.

------, L. H. Smith, G. F. Smoot, L. W. Alvarez, and M. A. Wahlig, 1972: Search for antimatter in primary cosmic rays. *Nature 236*, 335 - 338.

Ginzburg, V. L., and S. I. Syrovatskii, 1964: *The Origin of Cosmic Rays*. Pergamon Press, New York.

Pennypacker, C. R., G. F. Smoot, A. Buffington, R. A. Muller, and L. H. Smith, 1973: Measurement of geomagnetic cutoff rigidities and particle fluxes below geomagnetic cutoff near Palestine, Texas. *J. Geophys. Res. 78*, 1515 - 1527.

Smith, L. H., A. Buffington, G. F. Smoot, and L. W. Alvarez, 1973: A measurement of cosmic ray rigidity spectra for above 5 GV/c of elements from hydrogen to iron. *Astrophys. J. 180*, 987 - 1010.

run the data through our computer programs. The computer reconstructs the trajectory of the particle through the magnet for each event and determines the momentum and sign of the charge. Any events that appear to have negative charge are carefully remeasured by hand. Most of the "negative-charge" events are found to be the result of errors by the automatic measuring machine. Only a few hand-measured events are found still to curve in the negative direction, and their amount of curvature is insignificant.

We have recorded and analyzed almost 50,000 cosmic ray nuclei, but have identified none as antimatter. Perhaps our galaxy does not contain antimatter at all. If it does, the matter/antimatter barrier must be so strong that very few particles up to 0.999 the speed of light can get through. Our findings strongly limit the models that can be used by theorists who believe that antimatter exists in bulk in our galaxy. We would certainly like to improve our search, either by greatly increasing the statistics of the measurements, or by examining much higher energies. We seem to have reached the limit of our technique and are now searching for a better way to identify antinuclei. The beauty of this kind of search is that indisputable observations of only one or two antimatter events would be enough to assure us that antimatter exists in bulk in our galaxy. •

*Andrew Buffington is an associate research physicist at the Space Sciences Laboratory of the University of California at Berkeley, where he is involved in astrophysics experiments including flights of balloon gondolas to measure the nuclear, electron, and positron spectra of cosmic rays. He received B.S. (1961) and Ph.D. (1966) degrees from Massachusetts Institute of Technology. Before going to Berkeley in 1969, he worked as a research assistant at M.I.T., engaging in teaching as well as research in elementary particle physics. He has been a member of the Sigma Xi since 1961. His group is presently supported by a grant from NASA.*

# Long Duration Ballooning

*James A. Winker, Raven Industries, Inc.*

Most scientists who have used balloons as carrier vehicles for their experiments are aware that there are several types of balloons. The expandable neoprene type widely used in meteorological sounding typically rises until it bursts in a total flight time not greatly exceeding an hour. The so-called zero pressure balloon has accounted for by far the greatest proportion of scientific ballooning in the last quarter of a century. It rises to and floats at an altitude where it becomes full, attains buoyant equilibrium, and vents excess gas. Environmental changes or diffusion of the lifting gas disturbs the equilibrium which must be restored by an intermittent release of ballast. Ballasting requirements usually prohibit flights longer than three or four days; longer flights are generally undertaken during the biannual stratospheric wind reversal when it is easier and less expensive to track the balloon and recover its payload because it does not drift so far.

The primary requirement for achieving flight durations of great length is a carrier balloon. It is interesting to consider that as long duration ballooning becomes more common, a whole new philosophy will need to evolve to guide the conduct of flight operations and the design of scientific payloads. The balloon that will open this new era is called the superpressure balloon (Fig. 1) and is now in advanced stages of development; when perfected, it will increase flight durations by several orders of magnitude. The superpressure balloon has the inherent capability of long duration flight without ballast, as has been thoroughly demonstrated in NCAR's GHOST Program and subsequent exercises. Until about a year ago, the superpressure balloon that had evolved was mainly suited to atmospheric sounding and wind tracing. Now, however, the technology of super-

# 30- to 100-keV Protons Upstream From the Earth's Bow Shock

A74-21676

R. P. LIN AND C.-I. MENG

*Space Sciences Laboratory, University of California*
*Berkeley, California 94720*

K. A. ANDERSON

*Physics Department and Space Sciences Laboratory, University of California*
*Berkeley, California 94720*

Protons of 30 to 100 keV are found upstream from the bow shock whenever interplanetary magnetic fields connect the spacecraft and bow shock. Their energy spectrum is closely power law, $dJ/dE \propto E^{-\gamma}$ with $\gamma$ usually close to 3. The spectrum is sharply cut off above 100 keV. The protons do not appear upstream of a boundary determined by the solar wind speed (the speed at which the interplanetary field is being convected) and an effective upstreaming velocity of 2.5 to 3 times the solar wind speed along the field lines. We conclude that the protons are not accelerated locally at the bow shock nor are they accelerated in a Fermi process between the bow shock and the moving interplanetary magnetic field irregularities. We believe that the lower-energy (3–4 keV) protons accelerated and reflected by the bow shock and the Alfvén waves observed upstream are in some way responsible for the origin of the 30- to 100-keV protons in very large regions upstream from the bow shock.

In 1968, *Asbridge et al.* reported protons having energy a few times greater than the free-streaming solar wind protons flowed outward from the bow shock. The authors refer to this as transverse flow. The outward flowing proton density is 1–10% of the solar wind flow, and their total energy density at times may reach 50% of the solar wind energy density. Transverse flow protons are not found in the magnetosheath. The authors concluded that incoming solar wind protons were energized and reflected outward at the bow shock. The outward flows were frequently observed, and the authors concluded that they were a persistent feature of the bow shock. *Scarf et al.* [1970] reported measurements on what evidently is the same population of particles and referred to them as suprathermal protons. They found these particles to be correlated with electrostatic waves at frequencies near 3 kHz near the bow shock. These authors found in at least one case that the outward-flowing protons were associated with the presence of large amplitude electromagnetic waves. The frequent occurrence of waves in the 0.01- to 0.05-Hz range at times when the interplanetary field line connected to the bow shock had been known from magnetometer observations [*Heppner et al.*, 1967; *Greenstadt et al.*, 1968; *Fairfield*, 1969; *Russell et al.*, 1971].

*Fairfield* [1969] has suggested that since these Alfvén waves could not move upstream against the solar wind flow they must be locally produced in the solar wind and that it was the outward flowing protons moving against the solar wind that generated these waves.

In this report we present observations of protons upstream from the bow shock in the energy range 29–100 keV. These protons are found to exist continuously on interplanetary field lines that connect to the bow shock and may be related to the lower-energy protons and the waves that appear on these same field lines.

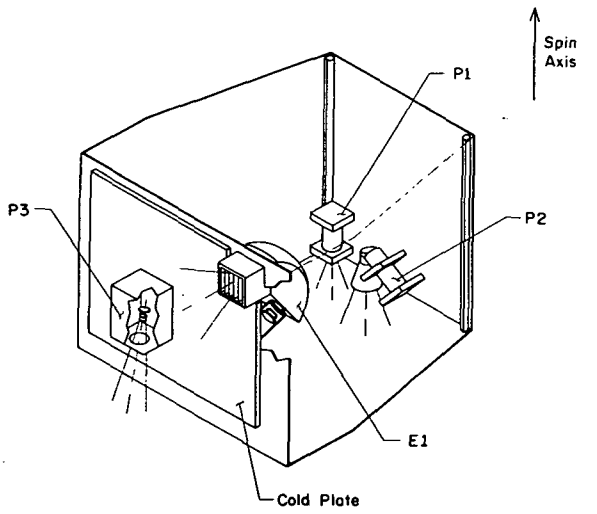
## EXPERIMENTAL DETAILS

The proton measurements used in this study were made aboard the Explorer 43 (Imp 6) spacecraft and the Apollo 16 subsatellite. The University of California experiment aboard Explorer 43 consists of two semiconductor detector telescopes and two Geiger-Müller detectors. Figure 1 shows the experiment package. The two semiconductor detector telescopes are mounted on a thermally isolated cold plate in front of the package. This plate is covered with a special coating to provide a large emissivity to absorptivity ratio. Although the plate views the sun for a portion of each spin of the spacecraft, the equilibrium temperature of the plate ranged from −68°C at launch to −55° a year later. The surface barrier detectors used in the telescopes have highest energy resolution and threshold noise characteristics below about −20°C. Table 1 summarizes the characteristics of the detectors.

Telescope P3 opens directly into space, has no foil covering, and consists of three solid state detectors: the first measures the energy of stopping particles, the second element forms a telescope with the first, and the third provides anticoincidence against penetrating particles. The axis of telescope is pointed to within 10° of the spin axis and is carefully collimated to eliminate scattered sunlight. This telescope measures both electrons from 18 to ~500 keV and protons from ~29 keV to 2 MeV. In order to take advantage of the high inherent energy resolution of the cooled surface barrier detectors in the P3 assembly, the output is pulse height analyzed into 64 channels through the on board spacecraft computer. Each channel is about 8 keV wide. This is approximately the full width at half-maximum energy resolution of the detectors for electrons and protons.

The Geiger-Müller tube P1 has a very thin (0.5 mg/cm²) mica window that admits electrons of >18 keV and protons of >200 keV. Protons below ~200 keV can be identified, since they are counted by P3 but not by P1. On March 13, 1971, at 1700 UT, Imp 6 was launched into a high apogee equatorial orbit from Cape Kennedy. Initial orbit parameters are: apogee, 211, 251 km; perigee, 6729 km; the orbital period is about 99½ hours. This spacecraft spends most of its time outside the earth's magnetosphere and bow

Copyright © 1974 by the American Geophysical Union.




1. This figure shows the Imp 6 experiment from which many of the upstream proton measurements were obtained. P3 is a telescope consisting of three cooled solid state detectors, P1 is a very thin window Geiger-Müller tube that responds to electrons to energies above 18 keV, and E1 and P2 are detectors that respond only to electrons.

shock but also passes through all regions of terrestrial trapped radiation from the distant radiation zones to the inner radiation belts.

The particle measurements on the Apollo 16 subsatellite come from two matched semiconductor detector telescopes, one of which is covered by a thin low $Z$ (parylene) foil. The foil stops protons below 344 keV, and the energy thresholds are matched between the two telescopes so that their response to electrons is identical. Thus the difference in counting rate between the open telescope and the foil telescope is due to protons below 344 keV. Threshold energy for the detection of protons is 38 keV. Further information on the detector complement for the Apollo subsatellites is given in *Anderson et al.* [1972]. The subsatellite was placed in orbit around the moon by the Apollo 16 astronauts on April 24, 1972, at an orbital height of about 100 km above the surface. Thus the subsatellite provides observations at a distance of about 60 $R_E$ from the earth.

## OBSERVATIONS

Protons of 30- to 100-keV energy are found on each orbit of Imp 6 that extends outside the bow shock. As we will show below, these protons are associated with the presence of the earth's bow shock. Figure 2 shows one orbit of Geiger tube P1 and telescope P3 data. Since the two detectors have equivalent 18-keV thresholds for electrons but quite different thresholds for protons (29 keV for P3 and ~200 keV for P1), the absence of response in P1 when P3 shows increased counting rate above background indicates that 29- to 200-keV protons are present in space. Electrons are observed predominantly inside of and close to the bow shock, with only short duration spikes ($\lesssim$5 min) upstream [*Anderson*, 1969], whereas >29-keV protons show large stable fluxes upstream lasting for as long as a few hours. Protons of 30 to 200 keV are also frequently found in the magnetosheath but are usually accompanied by electrons. Because of the complexity of separating out the proton fluxes in the presence of large electron fluxes, we have deferred discussion of the magnetosheath protons to a later paper. Figure 3 is a high time resolution (~5 s per readout) plot of an upstream proton event.

The characteristics of the upstream particle events are as follows:

1. Few electrons of >18 keV are found upstream from the bow shock compared with the large numbers of protons

TABLE 1. Imp 6 Detector Characteristics

| Detector Designation | Detector Type | Sensitivity: Electrons | Sensitivity: Protons | Geometric Factor, cm² sr | Look Angle FWHM,* deg | Angle to Spin Axis, deg | Channels |
|---|---|---|---|---|---|---|---|
| E1 | Surface barrier semiconductor detector in scatter geometry; anticoincidenced against penetrating particles. | 47–350 keV | None | 0.011 | 15 × 50 | 90 | 47- to 350-keV electrons in 16 sectors per spin<br>80- to 350-keV electrons in 4 sectors per spin |
| P1 | Open Geiger-Müller detector with 0.5-mg/cm² mica window. | >18 keV | >0.20 MeV | 0.027 | 35 | ~10 | ... |
| P2 | Geiger-Müller detector with 0.5-mg/cm² mica window in scatter geometry. | >20 keV | None | $2.5 \times 10^{-3}$ | 35 | ~10 | ... |
| P3 | Three-element telescope made up of surface barrier semiconductor detectors, one for anticoincidence against penetrating particles. | 18–450 keV | 29 keV 2 MeV | 0.10 | 37 | ~10 | 18- to 38.5-keV deposited energy<br>38.5- to 75.9-keV deposited energy<br>75.9- to 139-keV deposited energy<br>139- to 450-keV deposited energy<br>0.34- to 2.0-MeV protons<br>64-channel PHA†<br>18–450 keV |

* FWHM is full width at half maximum.
† PHA is pulse height analyzer.

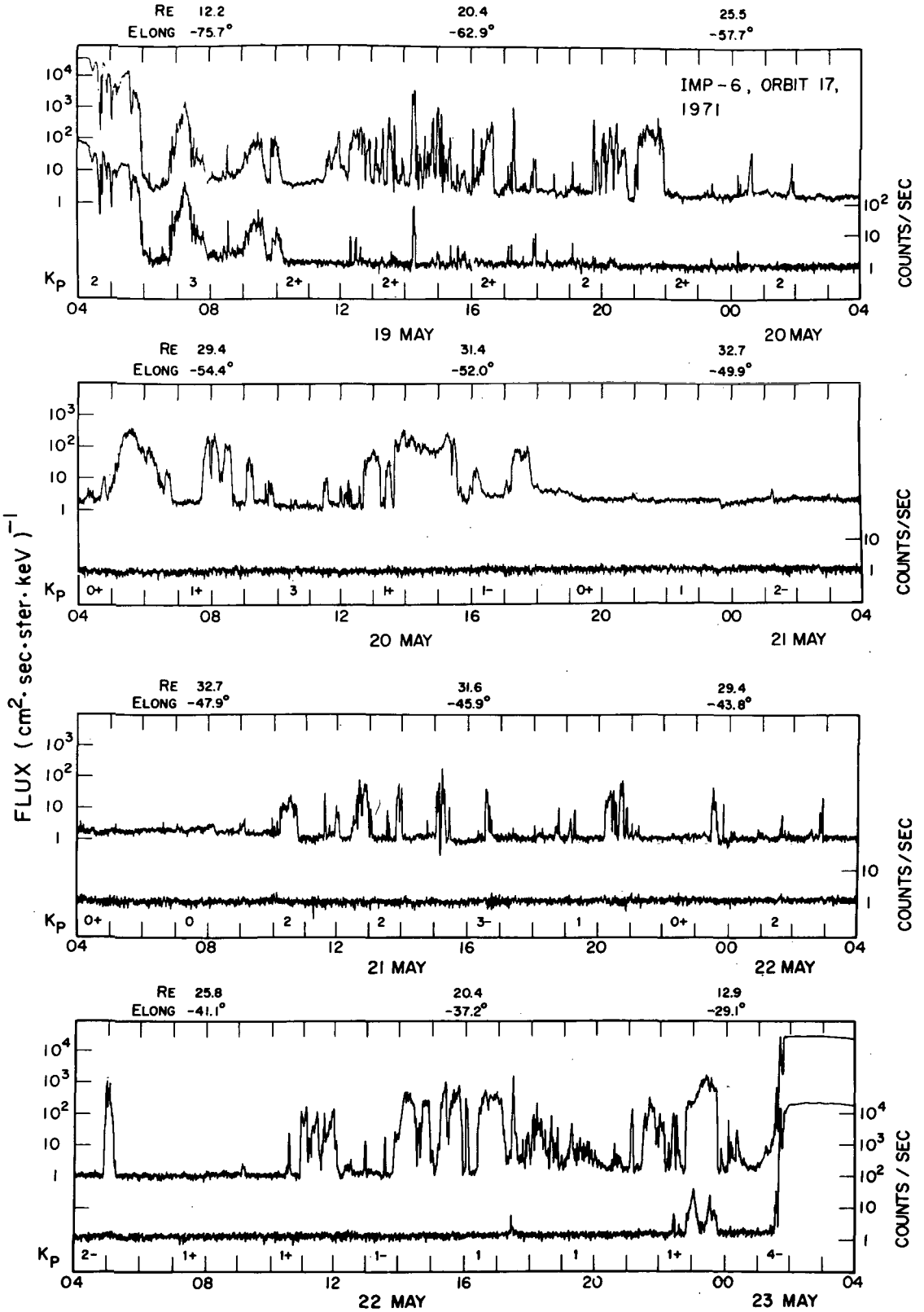



Fig. 2. Upstream protons of energy of 30–100 keV observed over a 3½-day interval. The top curve in each panel gives the rate of solid state detectors (P3). When there is response in P3 but no response in the Geiger tube (bottom curve) protons are present.

of $\gtrsim$30 keV. The electron fluxes occur in the form of short bursts lasting less than 5 min, whereas the proton fluxes have much longer duration. Inside the bow shock >.18-keV electrons are present as frequently as protons.

2. The upstream proton fluxes rise and fall in times of about 30 s or less. In this time interval the flux changes from very low background levels to proton intensities on the order of $10^3$ (cm² sr s keV)$^{-1}$ at 30 keV.

3. There is no striking or consistent asymmetry in the intensity versus time profile of the upstream protons.

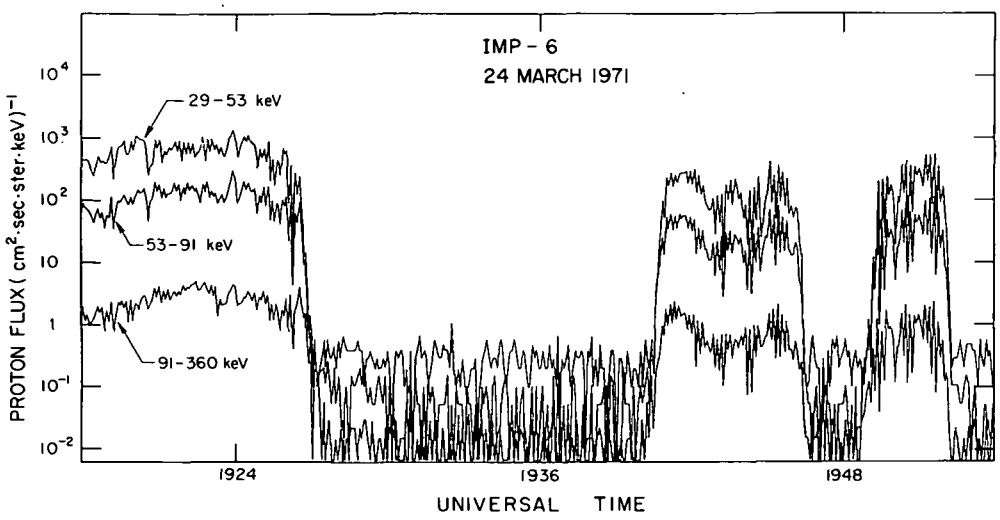



Fig. 3. Upstream proton fluxes are switched off and on when a field line convected to the bow shock passes through the spacecraft.

4. The duration of the upstream fluxes varies from a minute to several hours (Figures 4 and 5).

5. When the upstream protons appear for more than a few minutes, they attain a flux level that varies rather little from one event to another and only somewhat more from day to day. This level is in the range 1 to 10 × $10^2$ $(cm^2\ sr\ s\ keV)^{-1}$ and depends only slightly on the distance from the bow shock as will be discussed below.

6. Figure 6 shows where the Imp 6 spacecraft encountered upstream proton fluxes during two 4-month periods. The dawn side data were obtained from March through July 1971 and the dusk side from December 1971 through March 1972. These protons are encountered much more frequently when the spacecraft is on the dawn side of the earth's bow shock. This dawn-dusk asymmetry is confirmed by data obtained after March 1972.

7. Similar studies with the Apollo 16 subsatellite show the same result. Furthermore, using the direction of the magnetic field provided by Drs. P. J. Coleman and C. T. Russell, we find that when the magnetic field at the spacecraft is likely to intersect the bow shock, protons are almost always present. When the field at the spacecraft is unlikely to intersect the bow shock, the protons are almost never present.

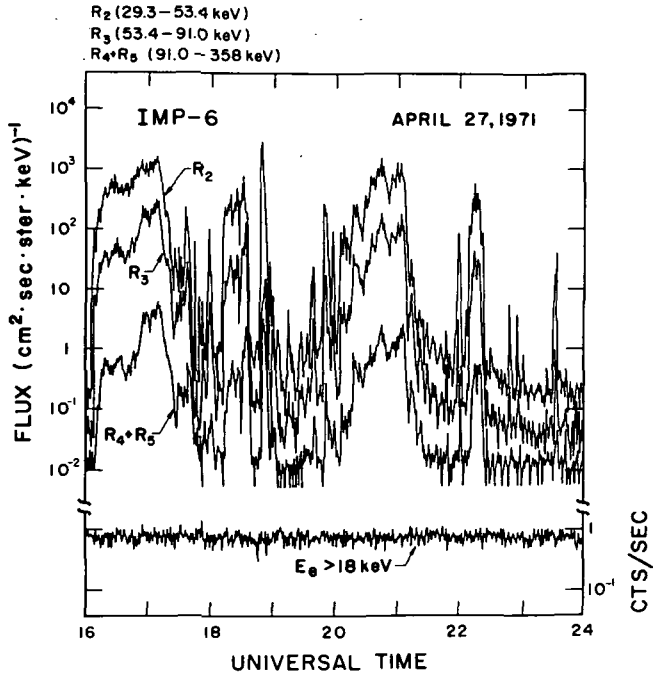



Fig. 4. Upstream proton fluxes showing the frequent and rapid variations due to the switching off and on by the magnetic field's directional variations. This figure shows that frequently during these times electrons above 18 keV are not present.

From the above observations we now draw two conclusions, which will be used throughout the remainder of the article.

The rapid increase in the proton intensity from background to the relatively stable flux level as seen, for example, in Figure 5, means that the spacecraft has at that moment come in contact with a field line connected to the bow shock. When a sudden decrease in intensity occurs, the spacecraft has become disconnected from the bow shock.

Fluxes of protons are continuously present on field lines that connect to the bow shock. The process that results in the acceleration of these protons is in a steady state. The near-constant level of intensity after switch on of these fluxes also points to the steady state nature of this process.

We now return to further observations made on the upstream protons.

## ANISOTROPIES

The Imp 6 P3 detector is oriented at an angle of 10° to the spin axis and measures particles within ~30° of the spin axis. For magnetic fields in the ecliptic plane, particles within ±30° of 90° pitch angle are measured as the spacecraft spins. In Figure 3, large regular modulations of the proton flux are apparent at all energies. These modulations may have peak to valley ratios of as high as 10:1, and their period is exactly the spin period. We have normalized these data to remove variations in the total flux that are not spin related. The degree of anisotropy and its direction are then found to change in time. Also the direction of the anisotropy is not aligned along the sun-earth line. Since the velocity of these protons is only 6–10 times the solar wind velocity, one might have expected a substantial anisotropy

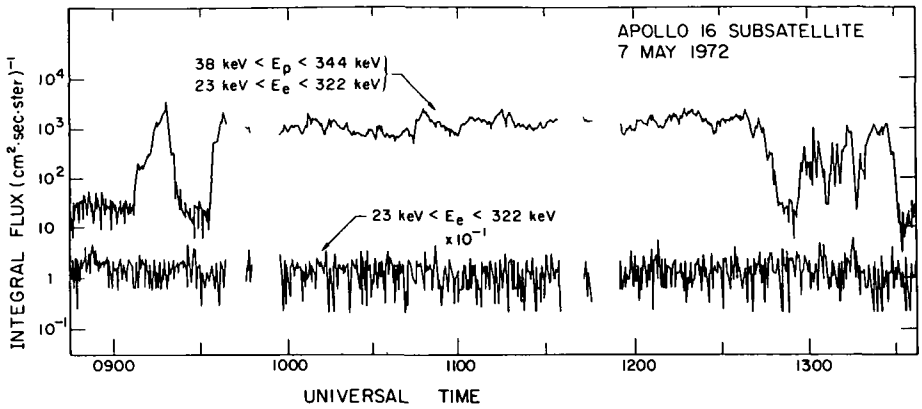



Fig. 5. This figure gives an example of a time when proton fluxes are continuously present upstream for over 3 hours. The near-constant level of intensity during this time is also striking and thus evidence for a steady state production of the protons is given.

along the sun-earth line due to the convection of these protons past the spacecraft by the solar wind.

However, the observations of the Apollo 16 subsatellite whose detectors are mounted on top of the spacecraft looking upward (normal to the ecliptic) appear not to observe substantial variations in the proton flux with variations in $\theta$, the angle the magnetic field makes with the ecliptic plane. Thus it may be that the protons are shadowed by the Imp 6 spacecraft itself in some unusual manner to produce this spin-associated variation. Further studies are needed to resolve the origin of this anisotropy.

## ENERGY SPECTRUM

The 64-channel pulse height analyzer provides high-energy resolution spectra of the P3 detector output. Several hundred spectra have been obtained over a several month period. Some typical spectra of the upstream protons are shown in Figure 7, and their features can be summarized as follows:

The energy spectrum from 30 to 100 keV is almost always well represented by a power law, $dJ/dE \propto E^{-\gamma}$ with $\gamma$ ranging from 2 to 6, but usually near 3. The spectra in Figure 7 are representative of the bulk of the spectra.

There is a very sharp break in the energy spectrum at about 100 keV. The 64-channel spectrum beyond that energy shows a $dJ/dE \propto E^{-9}$ or steeper spectrum when the quiet time background particle fluxes are subtracted. An $E^{-9}$ spectrum would be consistent with an absolute cutoff at 100 keV, since pulse pileup will result in some counts above 100 keV. However, some intense events have been observed in which protons are present with energies of $\gtrsim$200 keV, although with a very much steeper spectrum above 100 keV than below 100 keV.

We note that without the good energy resolution available from the Imp 6 experiment it would not be possible to show that the spectrum is indeed power law and we furthermore would not be able to determine that the spectrum has a very sharp cutoff at about 100 keV.

We have examined the energy spectra with best available time resolution at those moments when the proton fluxes are being switched on. We expect that for some fraction of the occasions the switch on would come about by the magnetic field 'pivoting' counterclockwise with respect to the spacecraft in such a way that the field line through the spacecraft suddenly makes contact with the bow shock somewhere near its nose. Then if the protons were accelerated at the bow shock, there would be a characteristic velocity dispersion of the protons moving along the interplanetary field line easily detectable by our experiments. We have never observed such energy dispersion, although the data we have to make this test is fairly limited. We cannot claim with certainty that we have ever encountered the situation described, although statistically it seems likely it has happened on some occasions. However, we consider it significant with respect to the origin of these protons that velocity dispersion has never been observed. We return to this point below.

Since our detectors do not distinguish between species of ions, we must consider the possibility that the particles observed are not protons at all but rather $Z \geq 2$ nuclei. A broad spectrum of $Z \geq 2$ nuclei is observed in the solar

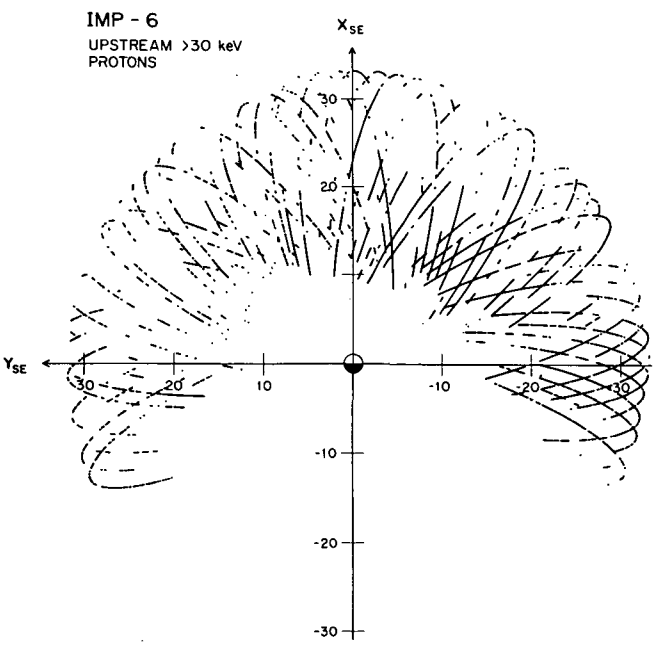



Fig. 6. The lines indicate when 30- to 100-keV protons were present during a total of 8 months of Imp 6 data. The dawn-dusk asymmetry is clear and thus indicates that the origin of the protons is associated with the presence of the earth's magnetosphere. The greater frequency of occurrence toward dawn is due to the preferred $\varphi = 135°$ or 315° direction of the interplanetary magnetic field.

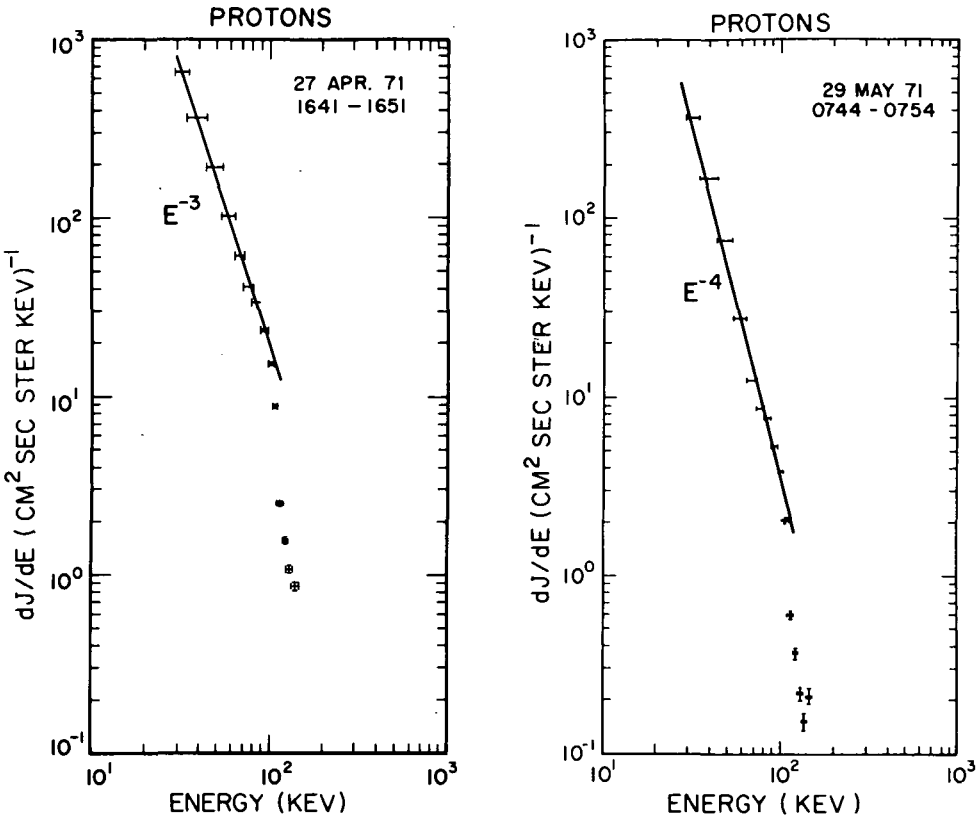



Fig. 7. This figure shows energy spectra of the upstream protons. The spectra are power law from 30 to 100 keV and are sharply cut off above 100 keV.

wind up to iron [*Bame et al.*, 1970]. If these nuclei are accelerated to the same velocity ($\sim 2.5V_{sw}$–$3V_{sw}$ or 6–9 keV/nucleon) as the low-energy protons of *Asbridge et al.* [1968], their energies would range from 24 to 36 keV for $He^4$ and from 340 to 500 keV for $Fe^{56}$. The pulse height observed by a surface barrier solid state detector for 500-keV iron nuclei would be about 100 keV. Even at 9 keV/nucleon, $He^4$ would not be observed, but $Z \geq 3$ nuclei would be observed. This explanation would be consistent with both the observed upstreaming velocity and the high-energy cutoff. However, the ratio of $Z \geq 3$ nuclei to protons that would be needed to produce the observed fluxes is about a factor of 50 higher than that in the solar wind. No evidence of such large enhancements of $Z \geq 3$ ions in the upstreaming particles has been reported.

It is possible to make a limited comparison of the upstream proton intensities in the range 30 to 100 keV observed in the present experiment to the intensity of the protons at lower energies observed in the Vela experiments [*Asbridge et al.*, 1968]. We do this by taking the typical power law spectrum at higher energies and extrapolating it to an energy of 3 keV. Clearly, this procedure can lead to only limited conclusions because different times and spatial regions are involved and there is always the possibility that two distinct particle species are involved whose spectra just happen to be similar. However, it is still of interest to do this extrapolation. From the Vela information we find that a representative value for the proton intensity at 3 keV is $10^6$ $(cm^2\ sr\ s\ keV)^{-1}$. When we extrapolate the spectrum shown in Figure 7 down to 3 keV, we get close to this value. Thus it is possible that the upstream proton population measured here is the continuation to higher energies of the Vela 'transverse flow.'

## Effective Velocity of Upstream Protons Along Interplanetary Field Lines

In order to quantitatively test the hypothesis that protons are accelerated somewhere on interplanetary field lines between the bow shock and the spacecraft, we have used observations from the Apollo 16 subsatellite, an example of which is shown in Figure 5. Figure 8 shows the location of Apollo 16 when the upstream proton events are observed. The magnetic field direction at the spacecraft will indicate whether or not the interplanetary field line intersects the shock. In addition, the interplanetary field lines are convected back by the solar wind so that the outer boundary of the protons depends on an effective upstreaming velocity $V_e$. This velocity $V_e$ could be the velocity of the protons themselves if the protons are accelerated in the shock and then move along the convecting interplanetary field lines, or $V_e$ could be the velocity with which the agent that accelerates the protons moves along the field line.

In the appendix we calculated the position of this boundary at 60 $R_E$ for various ratios of $V_{sw}/V_e$ and interplanetary field azimuth angles $\varphi$ for the average position of the bow shock. The period of observations was then divided into 2-min intervals, and those intervals with magnetic latitude $\theta > 10°$ were eliminated; i.e., only those values of $\theta$ likely to intersect the bow shock were retained. The presence or absence of the 30- to 100-keV protons for each 2-min interval was then plotted on an interplanetary field azimuthal direction versus spacecraft ecliptic longitude graph (Figure 9). Also shown on this graph is the location of the outer particle boundary at 60 $R_E$ for various values of $V_{sw}/V_e$. In Figure 10 the frequency of occurrence is plotted versus $V_{sw}/V_e$. The frequency of occurrence is >80% for times

when the interplanetary field intersects the bow shock, for an effective velocity $V_e$ of $2.5V_{sw}$ to $3V_{sw}$. It should be noted from Figure 9 that protons are occasionally observed even when the extrapolated interplanetary field line does not intersect the shock. These situations could arise if 'kinks' develop in the field line between the spaeecraft and the bow shock or if the bow shock position changes.

The above analysis of the frequency of occurrence was based on only a small period of observations and therefore must be considered preliminary until a more comprehensive analysis can be finished. However, we can draw the conclusion that >30-keV protons are essentially always present in upstream regions, behind a boundary that connects back to the bow shock, for an effective velocity of $2.5V_{sw}$ to $3V_{sw}$ along the interplanetary magnetic field lines.

### SPATIAL DEPENDENCE

Although the occurrence of protons clearly depends on ecliptic longitude, the attenuation with distance from the earth for these protons is remarkably small. Figure 11 graphs the maximum proton flux versus distance. If we include observations from Apollo 16, normalized to the same energy range as Imp 6, the total variation from 20 to 60 $R_E$ is only a factor of ~2 to 3. We have not included the <20-$R_E$ points, since they have a high likelihood of being inside the bow shock and thus contaminated with electrons. Typical fluxes for comparable levels of geomagnetic activity range from $\sim 5 \times 10^2$ to $1.5 \times 10^4$ $(cm^2\ s\ sr)^{-1}$ above 29 keV at 20- to 30-$R_E$ distance and from $\sim 5 \times 10^2$ to $10^4$ $(cm^2\ s\ sr)^{-1}$ at 60-$R_E$ distance. Figures 11 and 12 shows that there is an upper limit of 2 to $5 \times 10^3$ $(cm^2\ s\ sr\ keV)^{-1}$ at 30 keV, approximately equivalent to an upper limit total number density of >30-keV protons of $2.5 \times 10^{-5}$ $cm^{-3}$ and the upper limit energy density in these protons of $10^{-12}$ ergs $cm^{-3}$.

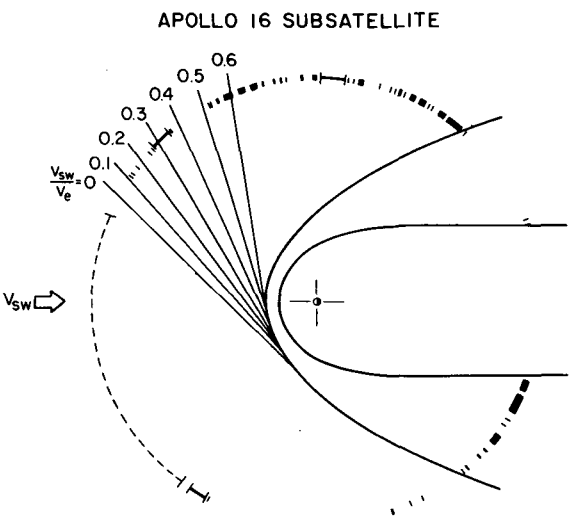



Fig. 8. The lines in this figure represent the boundary behind which the protons must lie for the case of interplanetary field direction $\varphi = 135°$ or 315° (see appendix). We find that only on a very few occasions do protons occur with $V_{sw}/V_e$ ratios less than about 0.3. The fact that the proton velocity gives $V_{sw}/V_e = 0.09$–$0.17$ is one reason we believe that the protons are not accelerated in the shock. We believe the acceleration is occurring throughout a large region of space owing to the presence of 2- to 5-keV protons and Alfvén waves. (Lines with end points indicate times when data are not available, dashed lines when intermittent data are available.)

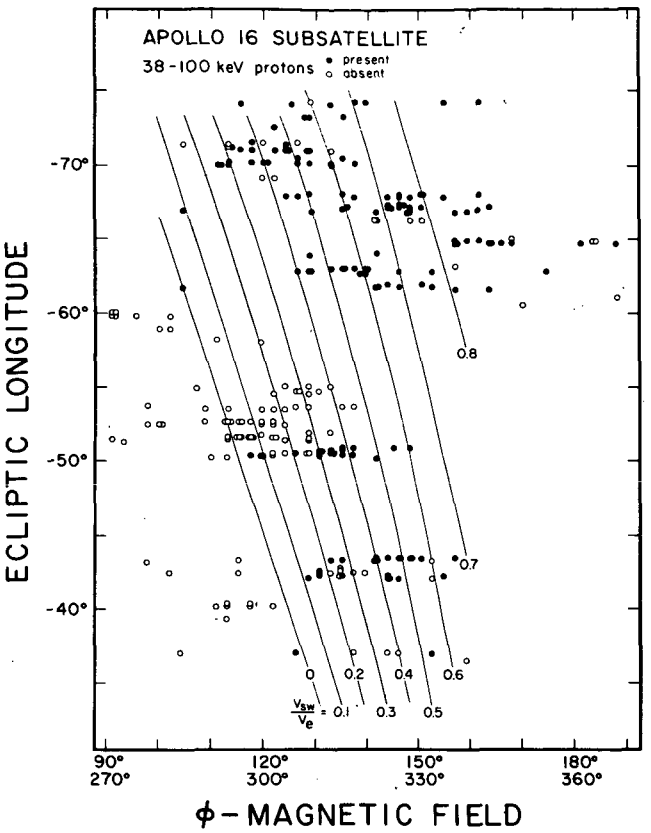



Fig. 9. This figure shows that for a wide range of interplanetary field azimuth angles the effective velocity is no greater than 4 or 5 times the solar wind speed. The solid and open dots indicate times when protons are present or not present, respectively, and the lines mark the location of the boundaries for various values of $V_{sw}/V_e$.

The upstream proton energy spectra also appear to vary with ecliptic longitude but not very much with radial distance. The dependence on ecliptic longitude is illustrated in Figure 13, in which every event large enough to provide a statistically significant determination of the spectrum in the 29- to 91-keV range is plotted versus ecliptic longitude. The median values of the ratio of counting rates R2 (29–53 keV)/R3 (53–91 keV) are indicated by the large symbols for each 30° longitude interval. Although a great deal of scatter is evident it is clear that (1) events near 0° ecliptic longitude are harder (smaller $\gamma$) than those at large ecliptic longitudes and (2) events at times of high geomagnetic activity are harder than those occurring during low activity.

The events appear to become progressively softer with ecliptic longitude toward the morning side, but no significant variation is evident going toward the dusk side except at the subsolar point where a sudden jump to softer spectra is observed.

We note here that for no events does the spectrum ever become harder than $dJ/dE \propto E^{-1.5}$. The importance of this finding is that there are no regions in which only protons of >53 keV are present without protons between 29 and 53 keV. We shall return to this point later.

### CORRELATION WITH *Kp*

In Figure 14 we plot the maximum level of proton intensity versus *Kp*. The covariance of these two quantities indicates that certain quantities or combination of quanti-

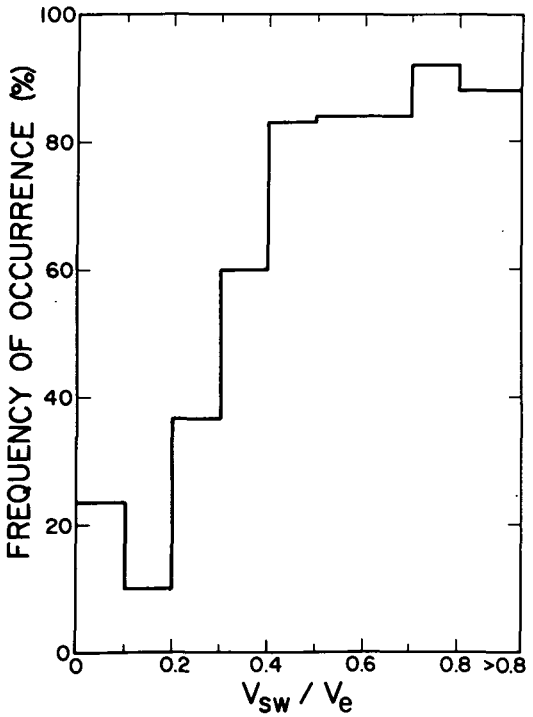



Fig. 10. This figure gives the results of Figures 8 and 9 in terms of percentage of the total observing times that protons are present for various values of $V_{sw}/V_e$.

ties of the interplanetary medium affect the proton acceleration process.

## DISCUSSION

We first summarize the two essential results.

Protons of 30 to 100 keV are always present in that part of the upstream interplanetary medium that is connected to the earth's bow shock by the interplanetary magnetic field. The upstream boundary of this region is characterized by an effective upstreaming velocity along **B** of 2.5 to 3 times the solar wind velocity. That is, the protons themselves or the agent responsible for accelerating them moves along the field lines at this speed while the field lines themselves are convected past the earth with the solar wind speed.

The switching on and off of the fluxes observed at the spacecraft arises from changes in the interplanetary field direction that connect or disconnect the field line passing through the spacecraft to the bow shock.

There are at least three possible origins for the energetic upstream protons: (1) They are produced at the bow shock, then travel upstream large distances (hundreds of earth radii or more); (2) They are produced by Fermi acceleration between the essentially static bow shock and magnetic irregularities in the moving interplanetary medium; and (3) They are produced throughout large regions of the upstream interplanetary region, but only on field lines connected to the bow shock.

For several reasons we do not favor the first possibility. First, the *Fairfield* [1969] 20- to 100-s period upstream waves have wavelengths of $8 \times 10^3$ to $4 \times 10^4$ km (if they are low frequency Alfvén waves being convected past the spacecraft). Fairfield reports that there is also enhanced power at slightly higher frequencies. Since the range of proton gyroradii for 29- to 100-keV protons is 5000 to $10^4$ km, there should be substantial scattering of these protons by upstream waves. This scattering should lead to a rapid falloff of the proton flux with distance from the bow shock. This is not observed. Second, the boundary of the proton region should be determined by the velocity of the 30–100 protons (6 to 13 times solar wind speed), even though our observations are predominantly of large pitch angle (~60°–90°) particles, because of the scattering in pitch angles. Third, if the 30- to 100-keV protons are produced at the bow shock and then run upstream along the field lines, some regions of the upstream medium could only be reached by the more energetic protons. For example, in Figure 8 the region between $V_{sw}/V_e = 0$ and $V_{sw}/V_e \approx 0.14$ could only be reached by $\gtrsim$50-keV protons. The proton energy spectrum measured in those regions would have a deficiency of low-energy protons. From Figure 13 we see that at no time is the proton energy spectrum (fit to $dJ/dE \propto E^{-\gamma}$) ever harder than $\gamma \approx 1.5$ and therefore a deficiency of low-energy particles is never observed.

A Fermi mechanism of the type proposed by *Axford and*

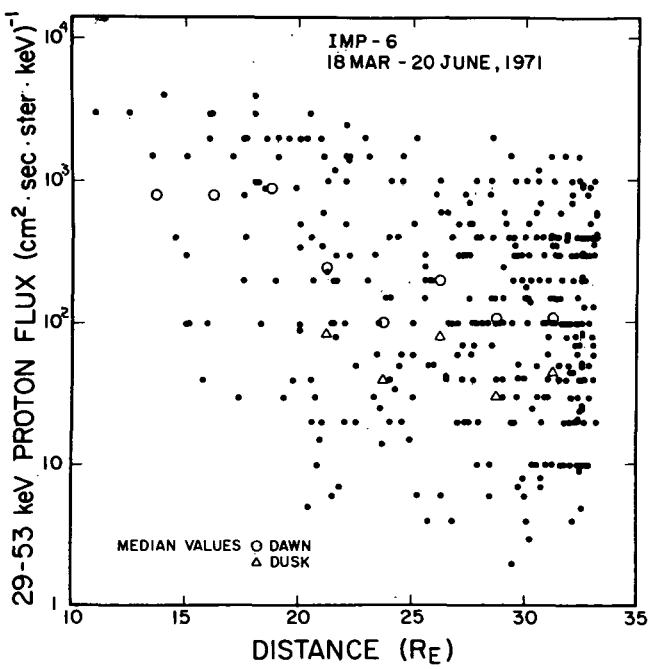



Fig. 11. The intensity of upstream protons changes slowly with distance from the bow shock.

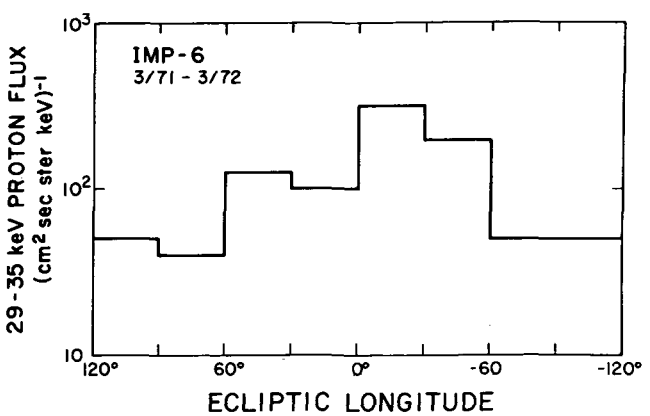



Fig. 12. The intensity of upstream protons is greatest when the field lines are more nearly normal to the bow shock surface near its nose.

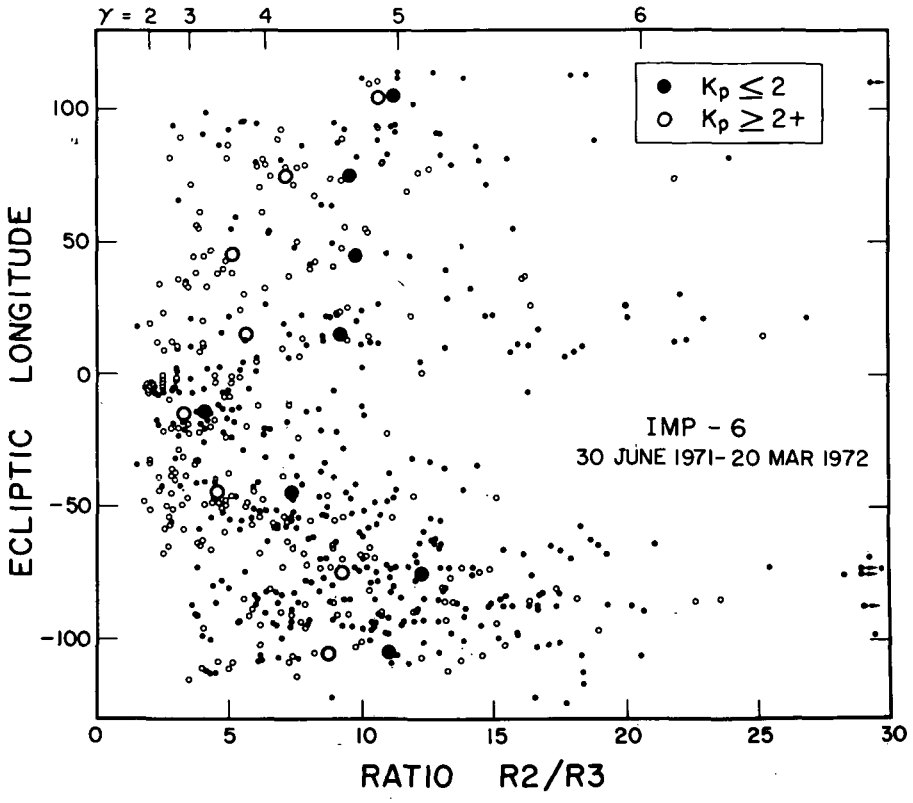



Fig. 13. The flattest spectra (smallest $\gamma$) occur at small ecliptic longitudes. Particles are present in this ecliptic longitude range when the interplanetary magnetic field lies close to the sun-earth line.

*Reid* [1963], operating by scattering particles between the bow shock and the wave-rich upstream interplanetary medium, may be very effective in accelerating protons. However, this mechanism should lead to a rapid falloff of the proton intensity with distance from the bow shock. This falloff is not observed. Additionally, the proton spectrum should harden from noon to dawn, since the energy gained will depend on the amount of time the interplanetary field line is connected to the bow shock. The opposite variation of the proton spectrum is actually observed (Figure 13). We therefore consider it most likely that the upstream protons are produced throughout the upstream region connected to the bow shock. In favor of that view we note that the effective upstreaming velocity derived here is the same as the velocity of the low-energy protons observed by *Asbridge et al.* [1968]. The upstream waves investigated by *Fairfield* [1969], *Russell et al.* [1971], and others are believed to be generated by these protons and to occupy the same spatial region as well.

The possibility of acceleration in the upstream inter-

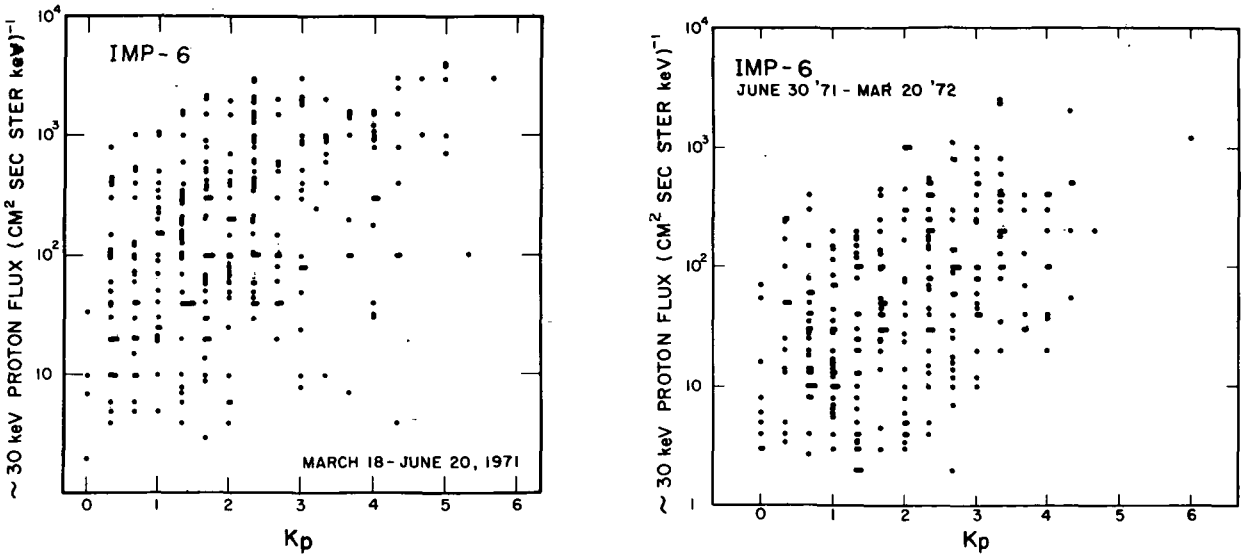



Fig. 14. The flux of upstream protons is quite strongly correlated with $Kp$ and thus indicates that the same parameters of the interplanetary medium that affect the magnetosphere as a whole in some way affect the proton acceleration process.

planetary medium is also favored by the observation that the upstream waves do not decrease rapidly with distance from the earth's bow shock (a factor of 2 over 15 $R_E$). The energy density of the solar wind is about $10^{-9}$ ergs/cm$^3$ compared with about $10^{-11}$ ergs/cm$^3$ for the upstream waves and $10^{-12}$ ergs/cm$^3$ for these 29- to 100-keV protons.

We plan further studies of the detailed correlations with lower-energy protons and waves.

## APPENDIX

The boundary of the region populated by particles on the interplanetary magnetic field lines upstream from the bow shock depends on the velocity of the particles or the accelerating agent and the direction of the magnetic field. In the ecliptic plane the average position of the bow shock can be closely approximated by a parabola (Figure 8)

$$x = a - (y^2/k)$$

where $x$ and $y$ are in solar ecliptic coordinates, $a$ is the distance from the center of earth to the subsolar point of the bow shock, taken here as 14 $R_E$, and $k$ is constant, taken here as 41 $R_E$ to give best fit to the observed average location of the shock.

Suppose the interplanetary field is in the ecliptic and is pointed at an azimuthal angle $\varphi$. Then a field line is described by

$$y = (x - V_{sw}t) \tan \varphi$$

where $V_{sw}$ is the solar wind speed and $t$ is time.

The intersection points of this field line with the bow shock are at

$$y = \frac{-k \cot \varphi \pm [k^2 \cot^2 \varphi - 4k(V_{sw}t - a)]^{1/2}}{2}$$

The intersection point nearest to or on the dawn flank is given by the minus sign in front of the square root. We shall only derive the dawn side boundary; the dusk boundary can be derived in a similar way.

Particles or the accelerating agents begin leaving the shock when their velocity exceeds the velocity of motion of the intersection point along the field line, i.e., when

$$V_e \sin \varphi = \frac{dy}{dt} = \frac{k V_{sw}}{[k^2 \cot^2 \varphi - 4k(V_{sw}t - a)]^{1/2}}$$

where $V_e$ is the effective velocity along the field lines. This occurs at

$$t_0 = \frac{1}{V_{sw}} \left[ a - \frac{k}{4} \left( \frac{V_{sw}}{V_e \sin \varphi} \right)^2 + \frac{k}{4} \cot^2 \varphi \right]$$

at a location

$$y_0 = \frac{k}{2 \sin \varphi} (\cos \varphi + V_{sw}/V_e)$$

$$x_0 = a - y_0^2/k$$

The boundary is defined by these first upstreaming particles. Their loci at later times are

$$y = y_0 - V_e(t - t_0) \sin \varphi$$

$$x = x_0 - V_e(t - t_0) \cos \varphi - V_{sw}(t - t_0)$$

These equations define a line $y = mx + b$ with slope

$$m = \frac{\sin \varphi}{\cos \varphi + V_{sw}/V_e}$$

and intercept

$$b = \frac{3k}{4m} - am$$

The boundary lines have been computed for various values of $V_{sw}/V_e$, and the lines for $\varphi = 135°$ are shown in Figure 8.

*Acknowledgments.* This work was supported by National Aeronautics and Space Administration contracts NAS 5-11038 and NAS 9-10509. We wish to thank P. J. Coleman, Jr., and C. T. Russell for making available magnetic field data from the Apollo 16 subsatellite.

* * *

The Editor thanks J. R. Asbridge and R. W. Fredericks for their assistance in evaluating this paper.

## REFERENCES

Anderson, K. A., Energetic electrons of terrestrial origin behind the bow shock and upstream in the solar wind, *J. Geophys. Res., 74,* 95, 1969.

Anderson, K. A., L. M. Chase, R. P. Lin, J. E. McCoy, and R. E. McGuire, Solar wind and interplanetary electron measurements on the Apollo 15 subsatellite, *J. Geophys. Res., 77,* 4611, 1972.

Asbridge, J. R., S. J. Bame, and I. B. Strong, Outward flow of protons from the earth's bow shock, *J. Geophys. Res., 73,* 5777, 1968.

Axford, W. I., and G. C. Reid, Increases in intensity of solar cosmic rays before sudden commencements of geomagnetic storms, *J. Geophys. Res., 68,* 1793, 1963.

Bame, S. J., J. R. Asbridge, A. J. Hundhausen, and M. D. Montgomery, Solar wind ions: $^{56}Fe^{+8}$ to $^{56}Fe^{+12}$, $^{28}Si^{+8}$, $^{28}Si^{+9}$, and $^{16}O^{+6}$, *J. Geophys. Res., 75,* 6360, 1970.

Fairfield, D. H., Bow shock associated waves observed in the far upstream interplanetary medium, *J. Geophys. Res., 74,* 3541, 1969.

Greenstadt, E. W., I. M. Green, G. T. Inouye, A. J. Hundhausen, S. J. Bame, and I. B. Strong, Correlated magnetic field and plasma observations of the earth's bow shock, *J. Geophys. Res., 73,* 51, 1968.

Heppner, J. P., M. Sugiura, T. L. Skillman, B. G. Ledley, and M. Campbell, Ogo-A magnetic field observations, *J. Geophys. Res., 72,* 5417, 1967.

Russell, C. T., D. D. Childers, and P. J. Coleman, Jr., Ogo 5 observations of upstream waves in the interplanetary medium: Discrete wave packets, *J. Geophys. Res., 76,* 845, 1971.

Scarf, F. L., R. W. Fredericks, L. A. Frank, C. T. Russell, P. J. Coleman, Jr., and M. Neugebauer, Direct correlations of large amplitude waves with suprathermal protons in the upstream polar wind, *J. Geophys. Res., 75,* 7316, 1970.

(Received July 19, 1973;
accepted October 19, 1973.)

A75-18016

# Suprathermal Particles

R. P. Lin

Space Sciences Laboratory
University of California
Berkeley, California 94720

Invited Review Paper Presented at the
Third Solar Wind Conference
Asilomar, Pacific Grove, California
March 25-29, 1974

## Abstract

Recent observations of suprathermal particles in the interplanetary medium in the energy range from solar wind ($\gtrsim$ 1 keV) to low energy solar cosmic rays ($\sim$1 MeV) are reviewed. A crude classification of the types of observed fluxes is given, and the sources of the suprathermal particle populations and energy contained in them are discussed. Among the possible interactions of these particles with the solar wind plasma, the particular interaction which generates type III solar radio bursts is analyzed.

## Introduction

In the past few years the first high energy resolution and high sensitivity measurements of suprathermal particles in the interplanetary medium have become available. Here we define the suprathermal energy range as between the solar wind ($\lesssim$ 1 keV) and previous solar cosmic ray measurements above $\sim$ 1 MeV. These new measurements show that (1) particles in this energy range are always present in the interplanetary medium; (2) their behavior is much more complex than at higher energies; (3) some of the particles interact with the solar wind plasma to produce waves; and (4) at times they contain the bulk of the total energy in the solar wind plasma. Here we briefly review the observations of suprathermal particles in the interplanetary medium.

## Observations

Although suprathermal particle fluxes exhibit highly complex behavior, some characteristic features observed at higher energies are also evident at these energies. These features provide a preliminary way of classifying the observed fluxes. They are: (1) flare particle events; (2) interplanetary shock associated spikes; (3) active region associated fluxes; and (4) "quiet" interplanetary fluxes. Additionally at energies of $\sim$ 5-50 keV, Frank (1970) reported large proton fluxes which appear to be associated with terrestrial magnetic storms.

Figure 1 shows six solar rotations of $\sim$ 30 keV particle fluxes (both electron and proton) from IMP-6. The largest transient increases are several flare associated events and some interplanetary shock events.

At energies above $\sim$ 10 MeV flare associated events are usually the only transient phenomena to be observed in protons. Protons have now been observed down to $\sim$ 40 keV from flares (Figure 2), and electrons down to $\lesssim$ 2 keV (Lin et al., 1974b). Electrons above $\sim$ 5 keV are frequently emitted by small flares or subflares (see Lin, 1974 for review) without a substantial accompanying proton flux (Figure 3). The frequency of e and p events is shown in Figure 4. Below $\sim$ 10 MeV interplanetary shock spikes are also in the proton flux (Palmeira et al., 1971). These events are very limited in extent ($\lesssim$ 1 hour) at high energies but highly complex

and accompanied by a broad pedestal at energies below ~ 1 MeV. Usually electrons above ~ 45 keV are not observed at interplanetary shocks but lower energy electrons are sometimes present as illustrated in a remarkable event observed by the Apollo 16 Subsatellite (Figure 5). These electrons could be the source of solar type II radio emission, which recently has been observed out to the orbit of the earth (Malitson et al., 1973).

During late May-early June, 1971 in Figure 1 active region particle fluxes possibly from McMath Plage #11340 are observed. Remnants of these fluxes may still be present in the next rotation.

Beneath the transient events a slowly varying background flux is present. The slowly varying background or "quiet" interplanetary fluxes in Figure 1 are composed of both electrons and ions.

Figures 6 and 7 summarize the energy spectra of various types of suprathermal particle fluxes as we know them now. One characteristic of non-terrestrial proton spectra is their power law exponent of approximately -1 below ~ 1 MeV. Since there are only a few spectra available in the energy range ~ 40 keV to 1 MeV, their commonality must be taken at present as a preliminary finding. That the spectrum of these protons extends to energies as low as ~ 50 keV places stringent limits on the amount of collisional energy loss suffered by these particles, if it is assumed that no other energy loss occurs in their propagation to 1 AU. For typical coronal temperatures of ~ $10^6$ °K the amount of coronal matter traversed by these particles is less than ~ 15 $\mu g/cm^2$. This would place the source of these particles at coronal densities of $\leq 10^9$ $cm^{-3}$. An alternative explanation for the flare particles might be that these particles go through the solar atmosphere in a fireball of a few times $10^7$ °K temperature. Collisional energy loss would be substantially reduced for both low energy protons and electrons in such a hot plasma (Anderson, 1972).

## Energetics and Sources of Suprathermal Particles

Because of the rising spectrum toward low energies the total energy density of these suprathermal particles, particularly protons, may at times dominate the solar wind plasma and magnetic field. However, in no case observed to date is the total energy density above ~ 40 keV more than ~ 0.1 of the bulk motional energy of the quiet solar wind, so that the general flow characteristics of the wind are unlikely to be substantially perturbed by these particles. In Table 1 the typical energy density for each class of suprathermal particle flux is listed along with the total energy or rate of energy input. Clearly large interplanetary shock and flare events may severely distort the quiet interplanetary medium.

The temporal association with the flare, and the velocity dispersion exhibited by the proton fluxes in Figure 2 down to energies of ~ 50 keV are strong evidence for the flare origin of those particles. What is not clear is whether these protons started at the sun with the energies they are observed to have at 1 AU. If the temporal profiles of the proton fluxes are interpreted as being due to the diffusion and convection of these particles through the interplanetary medium, substantial energy loss would be incurred through adiabatic deceleration in the outwardly moving and expanding solar wind. Gleeson (1971) estimates that factors of ~ 2.5-500 or more reduction in energy could occur for ~ 1 MeV proton (energy at 1 AU) in its propagation from the sun to the earth. On the other hand the large pitch angle anisotropies often observed for low energy protons suggests there is a lack of substantial scattering in the interplanetary medium and therefore not much energy loss for these protons. The amount of adiabatic deceleration in propagation of these protons remains a major unknown in the interpretation of suprathermal protons observed at 1 AU. In Table 1 the total energy estimate for the flare events is based on the assumption of no adiabatic deceleration and therefore is a lower limit to the true total energy.

Suprathermal protons observed in front of interplanetary shock may be the result of Fermi acceleration from reflections/scattering between the shock and upstream magnetic irregularities (Axford and Reid, 1963).

The interplanetary shock itself may be generated by suprathermal particles accelerated in the flare at the sun (Lin et al., 1974a). Figure 8 shows the August, 1972, series of large flares. What is striking is the one-to-one relationship between energetic proton injections by flares and interplanetary shock waves, and their relationship to the four largest hard (> 20 keV) X-ray bursts. An analysis of the largest and most energetic particle and shock event, that of 4 August, shows that the electrons which produce the observed hard X-ray burst through collisions with the solar atmosphere, can dump enough energy into the solar atmosphere through those collisions to account for the total energy of the interplanetary shock (assuming that the electron spectrum extends down to ~ 5 keV). The similarity of the proton spectra below ~ 1 MeV for flare events and interplanetary shock events suggests that flare protons of all energies may be accelerated by the shock in its passage through the solar corona. The magnetic fields present there are much stronger than those in the interplanetary medium and may thus lead to acceleration of the protons to higher energies.

The active region associated fluxes appear likely to be the result of continual acceleration by numerous small flares in the active region. McDonald and Van Hollenbecke (1973) showed that in some active regions many small importance 1 flares or subflares produced proton micro-events which, at energies above 10 MeV, were only observable by the highly

sensitive detectors used for galactic cosmic ray studies. In active region McMath plage #8905 these micro-events are sufficient to produce the observed active region fluxes at $\leq$ 1 MeV if the power law proton spectrum extended down to those energies.

The origin of the "quiet" background < 1 MeV proton fluxes (Krimigis, 1970; Lin et al., 1973b) is unknown at present. At energies below 1 MeV these particles are unlikely to be of galactic or near interstellar medium because of the difficulty of overcoming the modulation effects of the outward moving solar wind. The similarity of the 50 - 300 keV proton spectrum to the interplanetary shock and flare spectra suggests that these particles originate at the sun. Possibly the waves which are postulated to heat the corona may also be the source of these suprathermal protons.

The "quiet" interplanetary electrons below $\sim$ 100 keV (Lin et al., 1972) are also likely to be of solar origin, although we cannot rule out an interplanetary origin for these particles.

## Interactions With the Interplanetary Medium

Possible interactions of the suprathermal particles with the solar wind environment include: (1) The scattering of suprathermal particles by irregularities in the interplanetary magnetic field and consequent energy loss or gain; (2) the generation and damping of waves in the interplanetary medium by suprathermal particles, including (3) the generation of radio emission by suprathermal particles. Here we shall only describe in detail the results of studies on (3), in particular, recent work on the generation of type III radio bursts in the interplanetary medium.

Type III solar radio bursts are the most common type of impulsive phenomena observed from the sun. These bursts are characterized by a rapid frequency drift from high to low frequencies, and occasionally by the presence of two bands of emission, one at approximately twice the frequency of the other (see review by Wild, et al., 1963). A theoretical basis for the plasma hypothesis of Wild (1950) for type III solar radio bursts was first introduced by Ginzburg and Zheleznyakov (1958). A group of fast particles injected near the sun generate longitudinal electron plasma waves at frequencies near the local plasma frequency as they pass through the coronal plasma. These plasma waves then scatter off ion density fluctuations to produce electromagnetic radiation near the plasma frequency (fundamental), and off other plasma waves to produce emission at twice the plasma frequency (2nd harmonic). As the fast particles go upward in the corona and into the interplanetary medium the radio emission will drift from high to low frequencies. Typical drift rates for these bursts indicate velocities of $\sim$ 0.3-0.5 c for the particles where c is the speed of light.

The plasma waves are generated through a coherent Cerenkov plasma process. In order to produce plasma waves more rapidly than they are damped, the velocity distribution of the fast particles must have a positive slope, that is, a peak must exist in the non-thermal particle velocity distribution.

Recently Frank and Gurnett (1972) and Lin et al. (1973) have reported observations at 1 AU of energetic electrons and type III burst emission at the low frequencies characteristic of the near 1 AU plasma environment.

Lin et al. (1973a) located the position of the type III burst emission at each frequency (Figure 9) from the spin modulation of the radio signal. The emission originating at 1 AU was then compared with the particle data. Their results show (Figure 10) that the onset of the emission located at 1 AU corresponds to the arrival of electrons of $\sim 100$ keV energy. The subsequent build up of the radiation corresponds to the arrival of lower energy electrons, until maximum is reached when $\sim 10$ keV electrons first arrive.

The evidence from radio studies indicate that the second harmonic emission rather than fundamental is predominant for low frequency type III radiation (Fainberg et al., 1972; Malitson et al., 1973; Haddock and Alvarez, 1973). Thus, the frequencies of the near 1 AU radiation are $\sim$ twice the local plasma frequency. The observations of Frank and Gurnett would be in close agreement with Lin et al. (1973a) under the assumption of second harmonic radiation.

The peaked distribution needed to generate coherent Cerenkov radiation is clearly present in the electrons (Figure 11). This peaked distribution arises from velocity dispersion.

It should be noted that the electrostatic waves generated by the coherent Cerenkov processes, at least in these observations, are not so intense as to substantially modify the peaked distribution. That distribution evolves essentially as would be anticipated from velocity dispersion alone. The implication is that the time scale for relaxation of the beam through the wave-particle interaction is long compared to the time scale for evolution of the distribution by velocity dispersion.

These preliminary studies indicate that a quantitative plasma theory for type III bursts can be obtained through further observations of this kind. Such observations and theory, besides illuminating problems of plasma physics otherwise impossible to study, will also be applicable for the quantitative interpretation not only of solar flare radio phenomena, but possibly of galactic and extragalactic radio emission as well.

Acknoweldgements

This research was supported in part by NASA Contract NAS 5-10509.

References

Anderson, K.A., 1972, Solar Phys., 27, 442.

Axford, W.I. and G.C. Reid, 1963, J. Geophys. Res., 68, 1793.

Fainberg, J., L.G. Evans and R.G. Stone, 1972, Science, 178, 743.

Frank, L.A., 1970, J. Geophys. Res., 75, 707.

Frank, L.A. and D.A. Gurnett, 1972, Solar Phys., 27, 446.

Ginzburg, V.L. and V.V. Zheleznyakov, 1958, Astron. Zh. 35, 694, (English translation Soviet Astron. AJ, 2, 653).

Gleeson, L.J., 1971, Astrophys. Space Sci., 10, 471.

Haddock, F.T. and H. Alvarez, 1973, Solar Phys.,29, 183.

Krimigis, S.M., 1970, Acta Phys. Acad. Sci. Hung., 29, Suppl. 2, 125.

Lin, R.P., 1974, to be published in Space Science Reviews.

Lin, R.P., K.A. Anderson and T.L. Cline, 1972, Phys. Rev. Lett., 29, 1035.

Lin, R. P., L.G. Evans and J. Fainberg, 1973a, Astrophys. Lett., 14, 191.

Lin, R.P., H.S. Hudson and L. W. Jones, 1974a, in preparation.

Lin, R.P., R.E. McGuire and K.A. Anderson, 1973b, Phys. Rev. Lett., 31, 1268.

Lin, R.P., R.E. McGuire and K.A. Anderson, 1974b, to be published in the Proceedings of the I.A.U. Symposium 57 on Coronal Disturbances.

Malitson, H.H., J. Fainberg, and R.G. Stone, 1973, Astrophys. J., 183, L35.

McDonald, F.B. and M.A. Van Hollenbecke, 1973, High Energy Phenomena on the Sun, Symposium Proceedings, Ed. R. Ramaty and R.G. Stone, p.404.

Palmeira, R.A.R., F.R. Allum and U.R. Rao, 1971, Solar Phys. 21, 204.

Wild, J.P., 1950, Australian J. Sci. Res., A2, 541.

Wild, J.P., S.F. Smerd and A.A. Weiss, 1963, Ann. Rev. Astron. and Astrophys. 1, 291.

## ENERGETICS

| | | Energy Density At 1 A.U. | Total Energy |
|---|---|---|---|
| > 50 keV Protons* | Flare Events | 1 to $10^2$ ev/cc | $10^{29}$ - $10^{32}$ ergs |
| | Shock Events | $10^2$ to 5 x $10^2$ | $> 10^{27}$ ergs |
| | Storm Events | $10^2$ to 3 x $10^2$ | $\geq 10^{27}$ ergs |
| | Active Region Streams | $10^{-1}$ to 10 | $\geq 10^{19}$ - $10^{20}$ ergs/sec |
| | Quiet Background | $10^{-2}$ to $10^{-1}$ | $\geq 10^{19}$ - $10^{20}$ ergs/sec |
| Solar Wind | Interplanetary Magnetic Field | $\sim 10^2$ | $\sim 10^{25}$ ergs/sec |
| | Solar Wind Plasma Thermal Energy | $\sim 10^2$ | $\sim 10^{25}$ ergs/sec |
| | Solar Wind Bulk Motion | $\sim 5 \times 10^3$ | $\sim 5 \times 10^{26}$ ergs/sec |

* Energy in > 2 keV electrons in typically $\leq 10^{-2}$ the energy in protons.

Figure Captions

1. Shown here are the ~ 30 keV particle fluxes (18-38 keV electrons and 29-53 keV protons) over six solar rotations in 1971 as measured by the University of California experiment on IMP-6. Fluxes of terrestrial origin have been removed.

2. The proton fluxes from the over-the-limb event of 1 September 1971 observed by the Apollo 16 Subsatellite. The flare event clearly extends down to ~ 50 keV.

3. An example of a small flare electron event preceeded by a hard X-ray burst at the sun.

4. Plotted here are the number of flare events per month for both protons and electrons.

5. This interplanetary shock particle event observed on the Apollo 16 Subsatellite shows a clear splitting of the proton and electron fluxes in front of and behind the shock, respectively.

6. A summary of the energy spectra observed for the various types of suprathermal proton fluxes. Shown also are the terrestrial proton fluxes (dashed lines) observed at the orbit of the moon.

7. A summary of the energy spectra of various electron fluxes.

8. A summary of the August, 1972, series of large solar and interplanetary events. Note in particular the close correlation between: (i) new injections of flare protons; (ii) intense hard X-ray bursts; and (iii) interplanetary shocks.

9. The electron event of 16 May 1971 and the locations of the corresponding radio emission at different frequencies as observed on IMP-6.

10. The intensity vs. time profiles at different frequencies as compared with the arrival time of the energetic electrons for the 16 May 1971 event. The best agreement is with the 55 kHz emission, which is also the frequency at which the radio emission appears to be coming from very near 1 AU.

11. Energy spectra of the energetic electrons for the 16 May 1971 event as a function of time. The higher energy electrons arrive at 1 AU first, thus giving rise to peaked velocity distributions of the type needed for production of the observed radio emission.

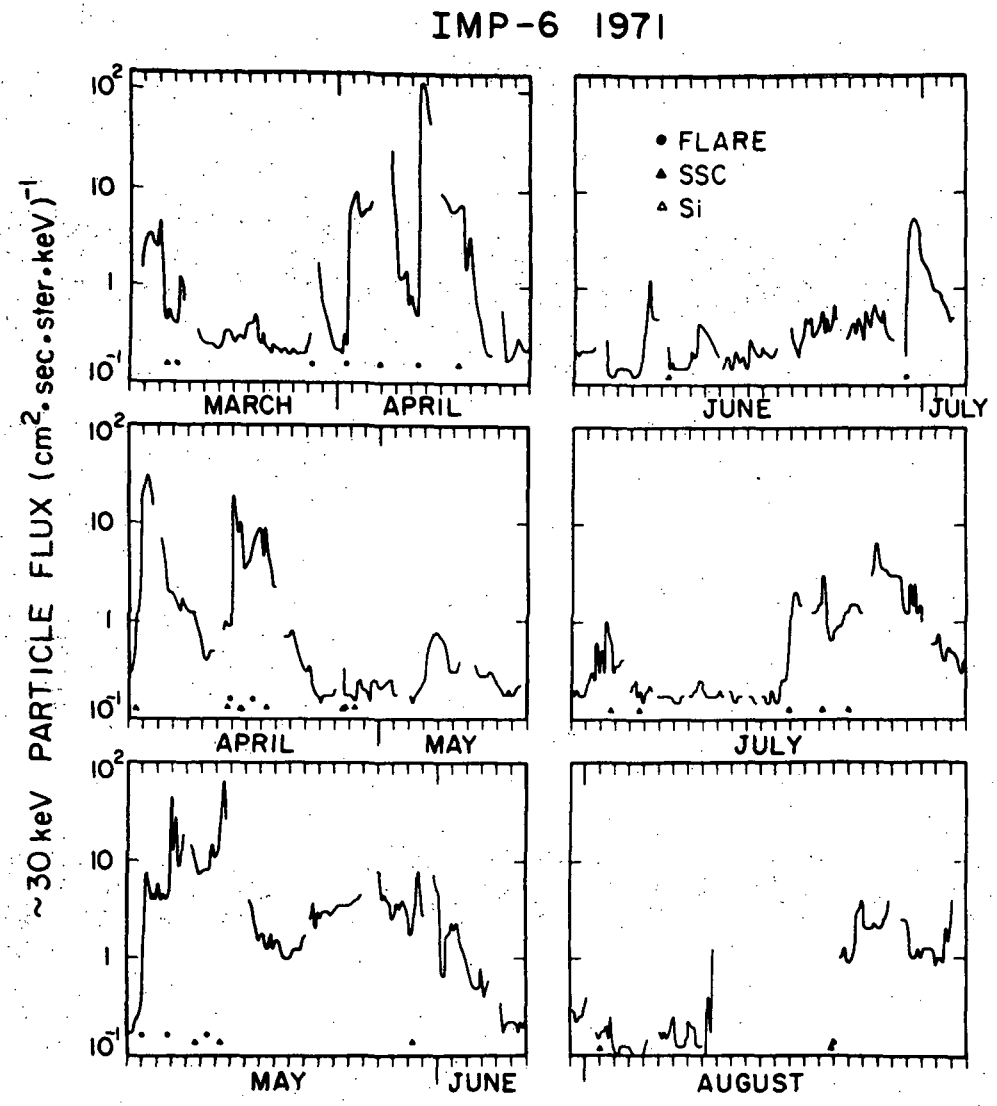



Figure 1

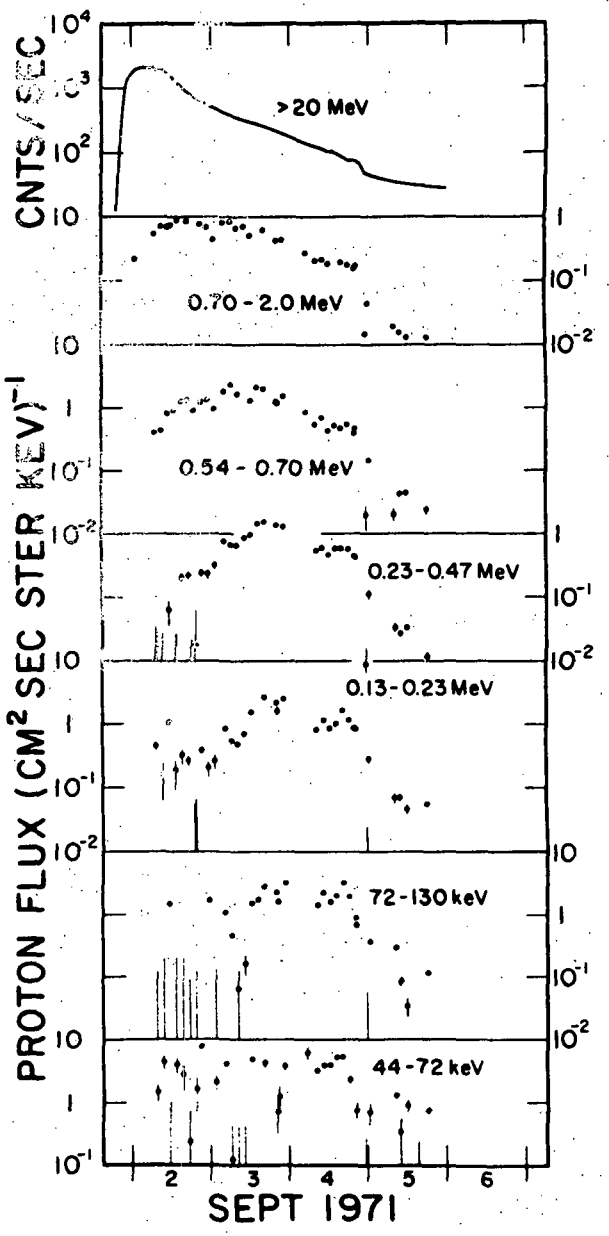



Figure 2

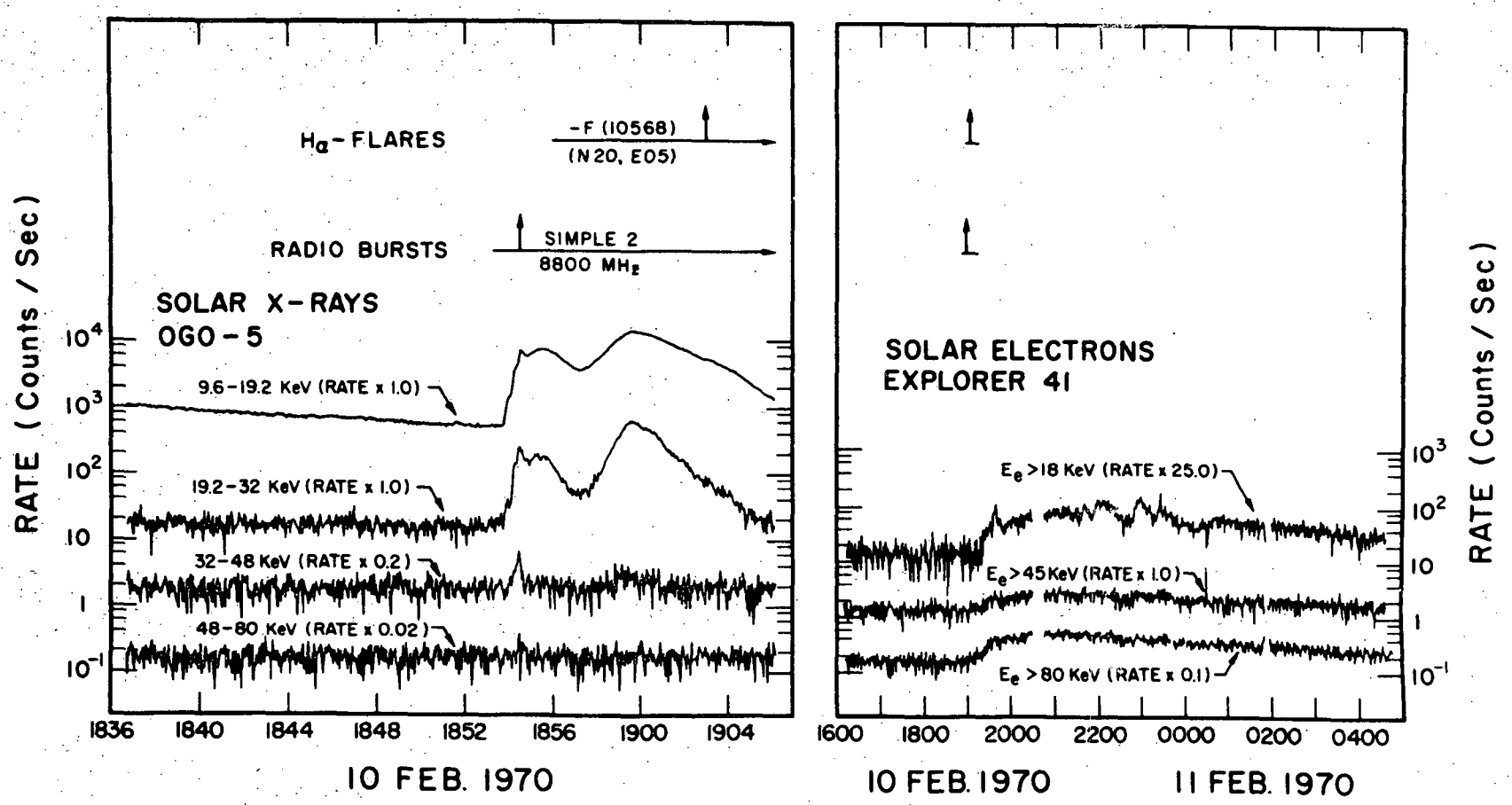

Hα-FLARES
-F (10568)
(N20, E05)
RADIO BURSTS
SIMPLE 2
8800 MHz
SOLAR X-RAYS
OGO-5
9.6-19.2 KeV (RATE x 1.0)
19.2-32 KeV (RATE x 1.0)
32-48 KeV (RATE x 0.2)
48-80 KeV (RATE x 0.02)
RATE (Counts / Sec)
10 FEB. 1970
SOLAR ELECTRONS
EXPLORER 41
Ee>18 KeV (RATE x 25.0)
Ee>45 KeV (RATE x 1.0)
Ee>80 KeV (RATE x 0.1)
10 FEB. 1970
11 FEB. 1970

Figure 3

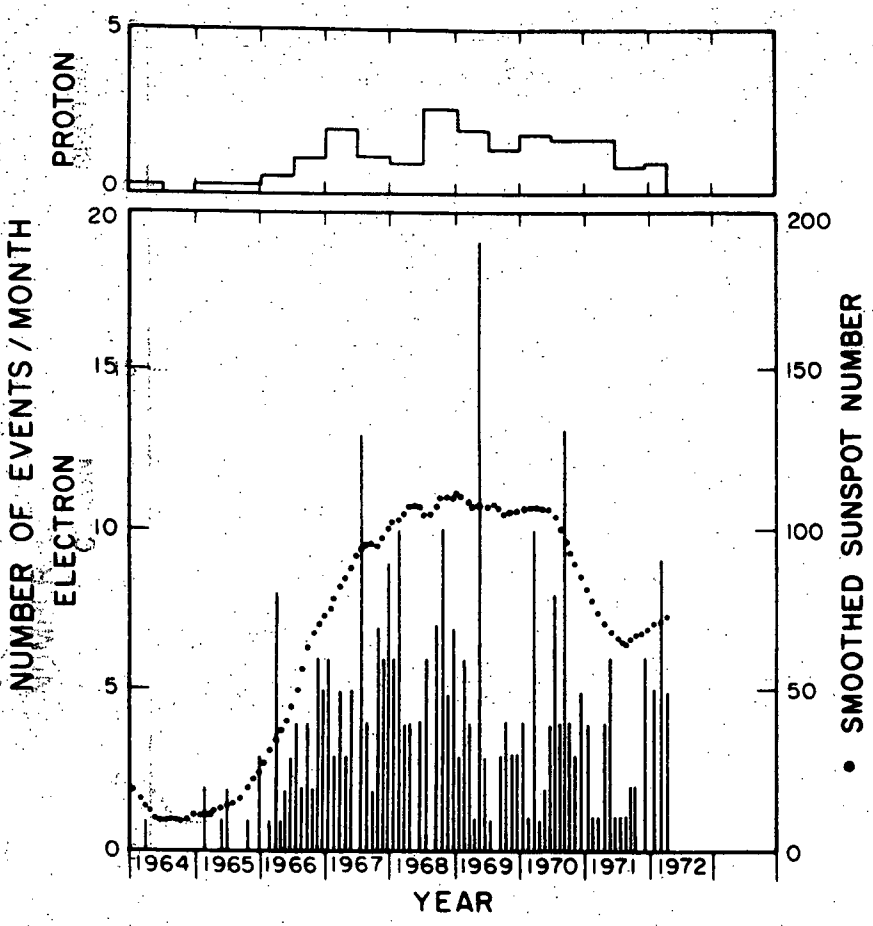



Figure 4

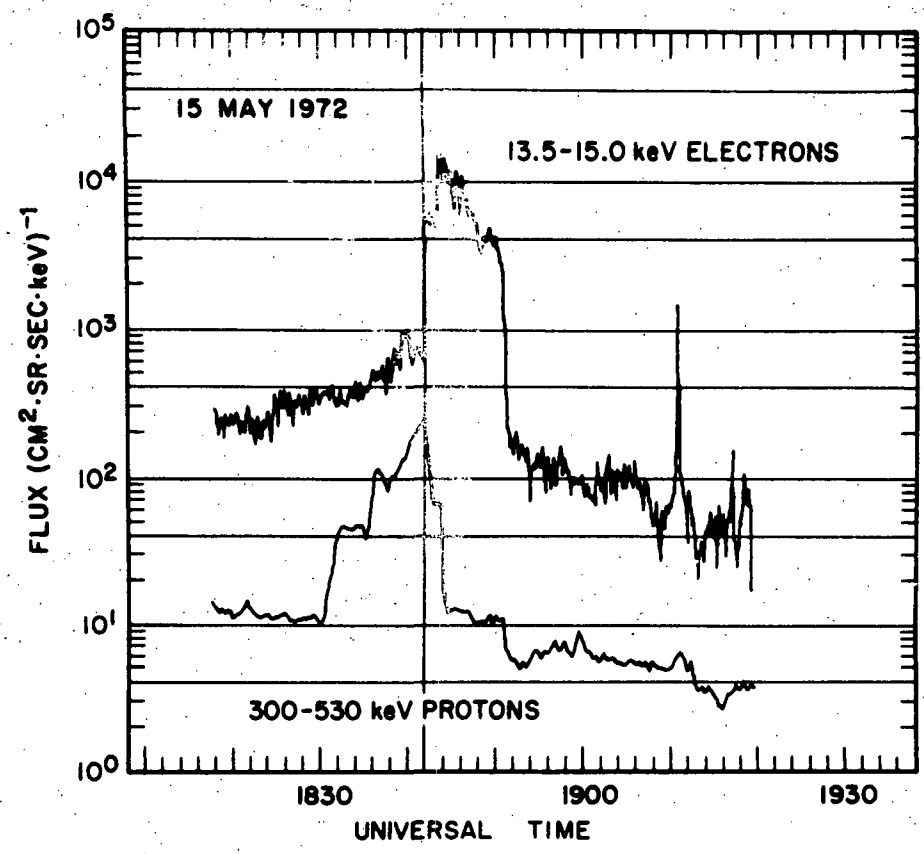



Figure 5

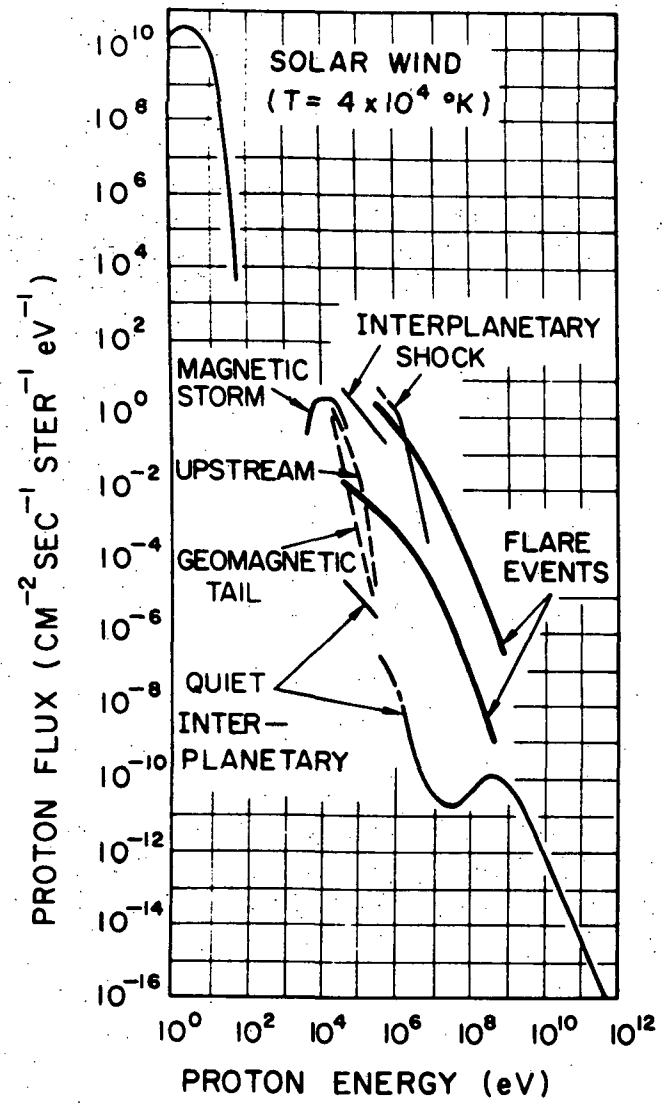



Figure 6

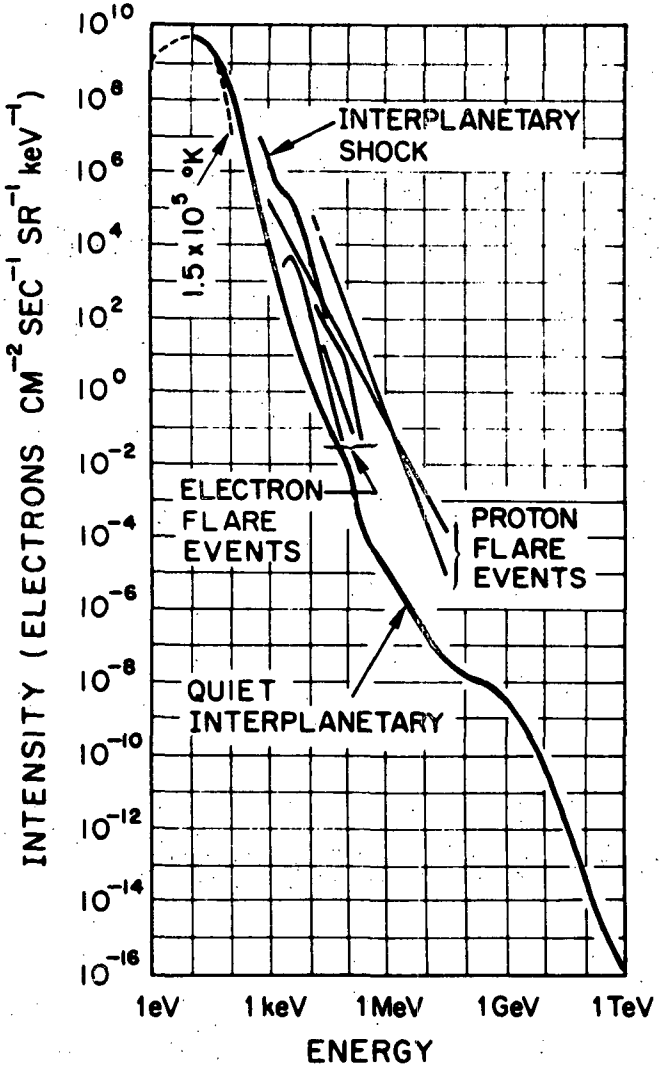



Figure 7

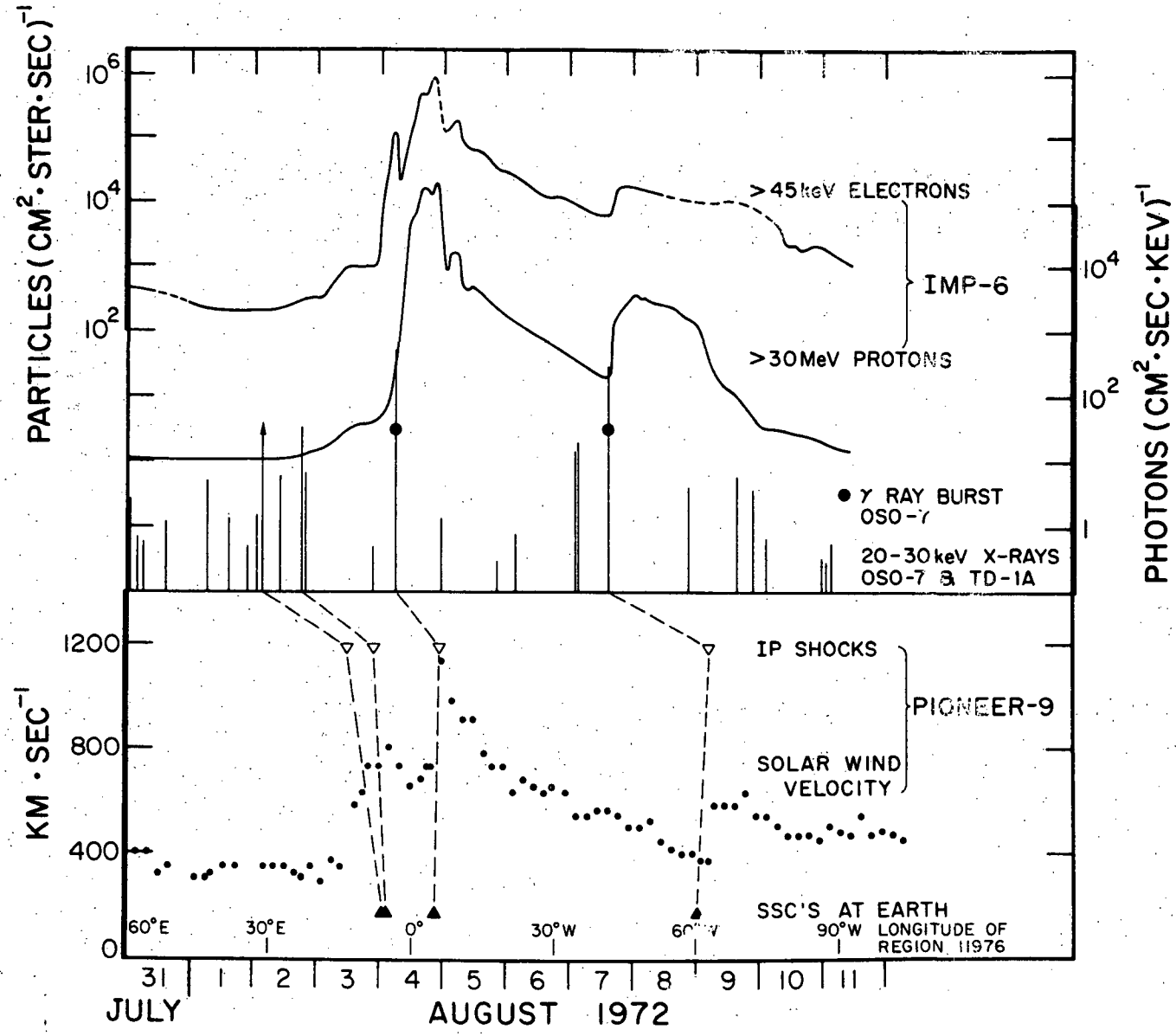
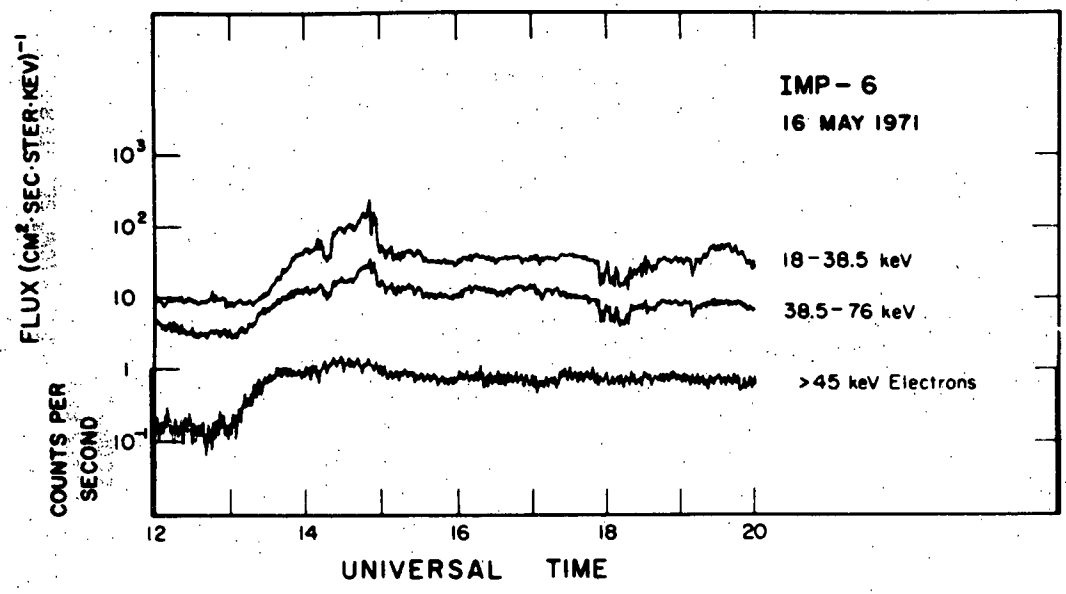

PARTICLES (CM²·STER·SEC)⁻¹
10⁶
10⁴
10²
>45keV ELECTRONS
IMP-6
>30MeV PROTONS
γ RAY BURST
OSO-7
20-30keV X-RAYS
OSO-7 & TD-1A
PHOTONS (CM²·SEC·KEV)⁻¹
10⁴
10²
1
KM·SEC⁻¹
1200
800
400
0
IP SHOCKS
PIONEER-9
SOLAR WIND
VELOCITY
SSC'S AT EARTH
60°E
30°E
0°
30°W
60°W
90°W
LONGITUDE OF
REGION 11976
31
1
2
3
4
5
6
7
8
9
10
11
JULY
AUGUST 1972

Figure 8


IMP - 6
16 MAY 1971
18 - 38.5 keV
38.5 - 76 keV
>45 keV Electrons
FLUX (CM² SEC·STER·KEV)⁻¹
COUNTS PER SECOND
10³
10²
10
1
10⁻¹
12
14
16
18
20
UNIVERSAL TIME

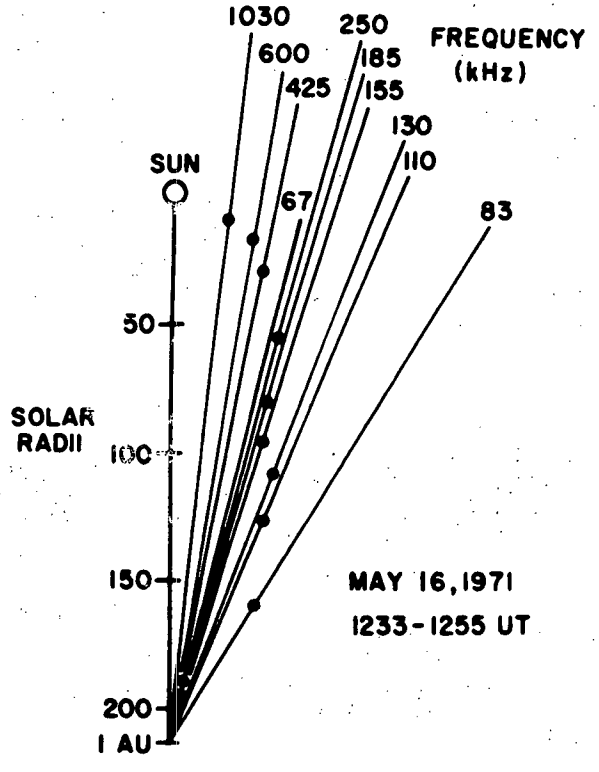

1030
600
425
250
185
155
130
110
83
67
FREQUENCY
(kHz)
SUN
SOLAR
RADII
50
100
150
200
1 AU
MAY 16, 1971
1233 - 1255 UT

Figure 9

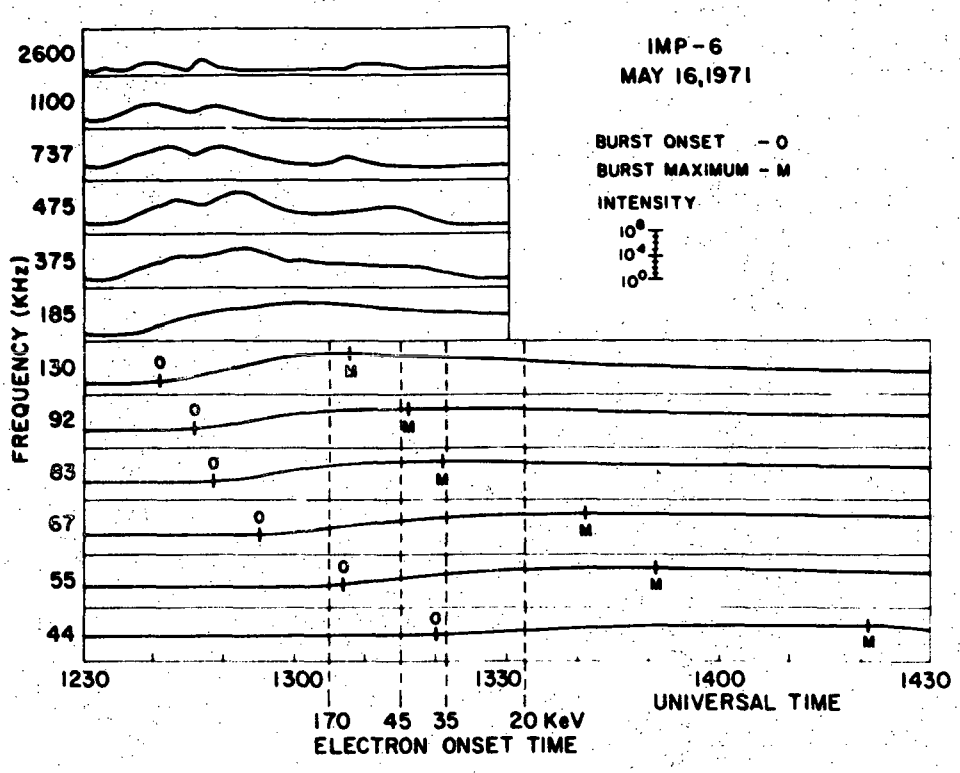



Figure 10

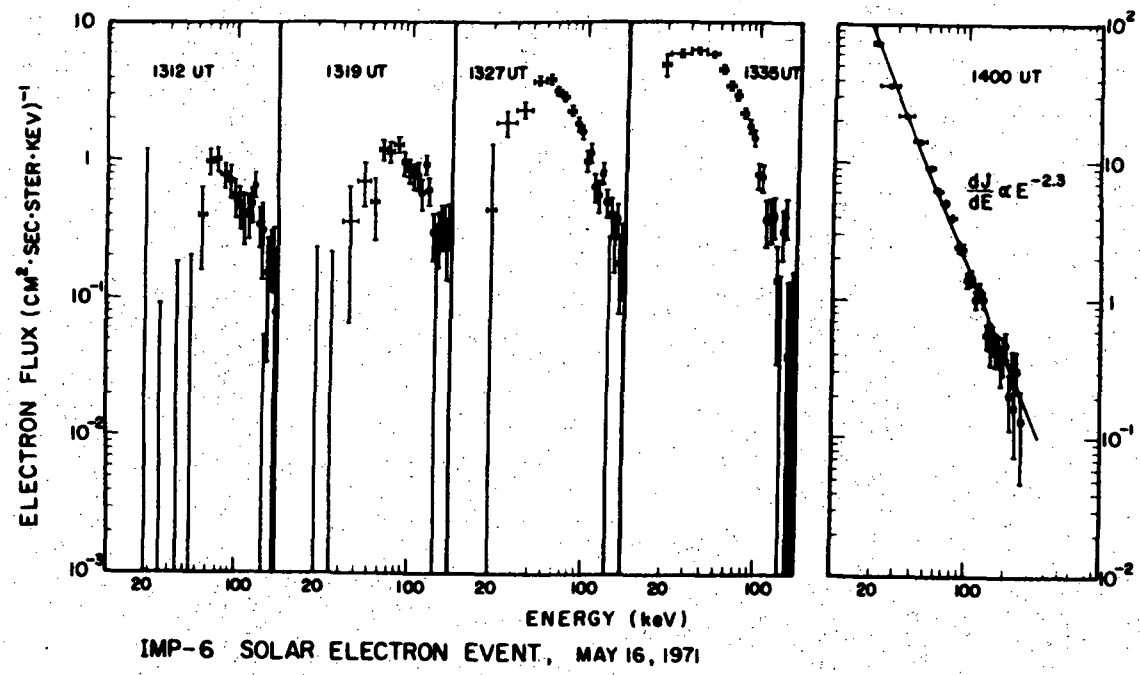



Figure 11

D2

# Magnetic Field Configuration in the Magnetotail near 60 $R_E$

by

C. I. Meng

and

K. A. Anderson

Space Sciences Laboratory
University of California
Berkeley, California 94720

March, 1974

Abstract

The magnetic field line configuration in the geomagnetic tail near 60 $R_E$ is examined on the basis of nearly three years of magnetometer data from the lunar orbiting Explorer 35 satellite. The magnetotail field line characteristics inside the plasma sheet and in the high latitude tail, i. e. outside the plasma sheet, are separately examined. In the high latitude tail the magnetic field lines systematically diverge in both longitudinal (east-west) and latitudinal (north-south) directions; the divergence angle between field lines near the dusk flank and those near the dawn flank is about $18^{\circ}$, but the divergence along the Z-direction is only about $8^{\circ}$. The degree of divergence is found to be greater in the evening side than that in the morning side, and greater in the northern than in the southern tail. Inside the plasma sheet, the field lines are greatly deviated away with respect to the sun-earth line and the spatial distribution of the field line deviation does not have any systematic tendency. The average longitudinal and latitudinal deviation is about $20^{\circ}$ and $17^{\circ}$, corresponding to the average $|B_Y|$ and $|B_Z|$ of 1.9$\gamma$ and 1.7$\gamma$, respectively. The geomagnetic activity dependence of the field line configuration is also examined and no drastic difference is found during different $K_p$ periods.

## I. Introduction

The magnetic field observations by IMP-1 satellite have revealed the basic structure of the magnetotail which consists of two oppositely directed lobes of magnetic field lines, one towards the earth in the northern tail and the other away from the earth in the southern tail (Ness, 1965). Between these two regions is the plasma sheet, a slab of plasma with a finite $\beta$ value. The mid-plane of the plasma sheet coincides approximately with the neutral sheet which is formed by a thin current sheet, separating the oppositely directed field line fluxes, to sustain the magnetotail. Characteristics of the cislunar magnetotail field have since been explored extensively by IMP and AIMP series of satellites. The magnetic field is usually very stable and oriented approximately along the Sun-Earth line; the aberration angle of the magnetotail axis, with respect to the Sun-Earth line is about $3.1^{\circ}$ due to the motion of the Earth around the Sun and/or the non-radial solar wind flow associated with the corotational motion of the interplanetary plasma with the Sun (Ness et al., 1967; Behannon, 1968, 1970; Behannon and Ness, 1968; Mihalov et al., 1968; Sonett et al., 1969). The divergence of magnetotail field lines away from the magnetotail axis was indicated by Explorer 33 and 35 magnetic field measurements (Mihalov et al., 1968; Behannon, 1970). However, it is recently reported that the field lines inside the plasma sheet are highly fluctuating and deviate considerably from the Sun-Earth line (Ness et al., 1967; Mihalov et al., 1968, 1970; Behannon, 1970; Hruska and Hruskova, 1970). Furthermore, the plasma sheet in the magnetotail undergoes frequently large-scale temporal variations, for

example during substorms. If this is indeed the case, it is important to attempt to separate spatial (with respect to the neutral sheet) and temporal variations.

In this paper, the configuration of the tail magnetic field lines both inside and outside of the plasma sheet at the lunar orbiting distance is separately examined by using nearly three years of magnetic field observations from the Explorer 35 satellite. In addition to the two-dimensional (YZ plane) mapping of the field line spatial distribution perpendicular to the magnetotail axis, temporal changes of the field line configuration are also studied for different geomagnetic activities represented by the $K_p$ index; seasonal variations are also investigated.

## II. Analysis

The once every synodical month magnetotail crossing by the lunar orbiting Explorer 35 provides the most extensive magnetic field measurement in the distant magnetotail. The magnetic field data used here are obtained from the on-board NASA Ames magnetometer, averaged over either 7 or 8 telemetry sequences (approximately every 10 minutes) and this set of data has been used and described in earlier studies of the plasma sheet geometry (Meng and Mihalov, 1972a, b). In order to avoid possible inclusion of the magnetic field observations outside the magnetotail in the magnetosheath region, measurements obtained within the range of $-21\ R_E \leq Y_{SM} \leq 24\ R_E$ and $-18\ R_E \leq Z_{SM} \leq 18\ R_E$ are used here. The asymmetry along the $Y_{SM}$ axis is due to the tail axis aberration. There are about 16,000 10-minute magnetic

field averages from thirty-four Explorer 35 magnetotail crossings between July 22, 1967 and March 24, 1970. (See Table 1 of Meng and Mihalov, 1972a, for detailed date of crossings.)

In the high latitude tail lobes the magnetic field lines align approximately with the Sun-Earth direction; whereas inside the plasma sheet, their direction is highly fluctuating and is not simply directed toward or away from the sun. A binary test was utilized to divide the magnetic measurements into these two different regions. Two criteria are used independently to label magnetic field observations in the high latitude magnetotail region (outside the plasma sheet): (1) when more than 90% of the individual vector samples, within the 10-minute interval, are directed parallel to the Sun-Earth line between $-13^{\circ} \leqslant \theta_{SM} \leqslant 13.0$ and $-25.0^{\circ} \leqslant \varphi_{SM} \leqslant 15.0^{\circ}$ (or $158.0^{\circ} \leqslant \varphi_{SM} \leqslant 195.0^{\circ}$); (2) when the variance over the 10-minute period, along the average magnetic field, is less than $0.8\gamma$. Otherwise, the observed magnetic field is defined as those detected inside the plasma sheet (Meng and Mihalov, 1972a). Results from both criteria are similar, thus only those using criterion (1) will be presented in this paper.

By ignoring the ellipticity of the moon's orbit and the variation of the Explorer 35 orbit along the Sun-Earth line (X axis), the magnetic measurements are "mapped" onto two dimensional (YZ) plane. The cross section of the magnetotail from $Y = +24\ R_E$ to $Y = -21\ R_E$ and from $Z = +18\ R_E$ to $Z = -18\ R_E$ was divided into 180 square areal elements with a $3\ R_E$ by $3\ R_E$ size. Figure 1 shows the distribution of the number of data points in each areal element, in which the

upper number is for those detected in the high latitude tail and the lower number for those in the plasma sheet. It is important to note that in this paper, the terms "high latitude tail" and "plasma sheet" refer to the two magnetotail plasma regimes and do not refer to the averaged geometrical locations in the magnetotail; thus, the field lines of "high latitude tail" can sometimes be detected at locations adjacent to the expected neutral sheet location ($Z_N = 0$ or $Z_{SM} = 0$) and similarly, the field lines of the "plasma sheet" can sometimes be observed far away from the equatorial region of the magnetotail.

## III. High latitude tail region

The total of about 1200 hours of high latitude tail magnetic field observations are evenly distributed either in the northern (+Z) versus southern (-Z) or the morning ($-21\ R_E < Y_{SM} < 3\ R_E$) versus evening ($3\ R_E < Y_{SM} < 24\ R_E$) parts of the magnetotail (see Figure 1). Within each areal element, averages of the following magnetic field components in the solar magnetospheric coordinates are obtained: (1) the field magnitude (B); (2) latitude angle ($\theta$) which is the north-south angle between the field vector and the solar magnetospheric equator; (3) longitude angle ($\varphi$) which is the east-west angle between the magnetic vector and the Sun-Earth line on a plane parallel to the solar magnetospheric equator; (4) X-component ($B_X$); (5) Y-component ($B_Y$); (6) Z-component ($B_Z$), which are three field components in the solar magnetospheric coordinates, together with their absolute values. Averages over each column and row are also calculated for examining the longitudinal (along $Y_{SM}$) and latitudinal (along $Z_{SM}$) distributions

of these field parameters. In the spatial distributions of these quantities on the cross section plane of the magnetotail, the geometrical location of each magnetic field measurement is expressed in the solar magnetospheric coordinates and two neutral sheet coordinates, defined by Murayama (1966) and Russell-Brody (1967). These three coordinate systems give essentially similar spatial distributions of the field parameters; therefore, only Murayama's neutral sheet coordinates are presented here.

A. Spatial distribution

The average of the magnetic field magnitude in the high latitude tail is 10.8$\gamma$ and the scatter with respect to this value is illustrated in Figure 2. Figure 3 shows that the distribution of the field magnitude (B) is not uniform over the tail cross section. Longitudinally, the magnitude in the central part of the tail ($-9\ R_E < Y_{SM} < 9\ R_E$) is higher than that near both flanks ($9\ R_E < Y_{SM} < -9\ R_E$), and , latitudinally, the field intensity in the southern tail is higher than that in the northern tail.

The spatial distribution of $B_X$ (the average of the field X-component) and $|B_X|$ (the average of the absolute value of $B_X$'s) at each areal element is illustrated in Figure 4, in which the $B_X$ (the upper value in each element) is positive in the positive Z-region and negative in the negative Z-region, except for a few irregularities, corresponding to a sunward and anti-sunward directed field in the northern and southern part of the tail, respectively. It is interesting to compare the value of $B_X$ with that of $|B_X|$ at each areal element;

a smaller value of $B_X$ indicates that the field with both polarities in X-direction were observed. For example at $6\ R_E < Y_{SM} < 9\ R_E$ and $3\ R_E > Z_N > 0\ R_E$, $B_X$ of -0.15γ and $|B_X|$ of 10.56γ means that approximately equal numbers of sunward and anti-sunward directed fields were observed here. Areal elements with different values of $B_X$ and $|B_X|$ are limited in regions above $Z_N = -3\ R_E$ and below $Z_N = +6\ R_E$, due to temporary changes of the neutral sheet configuration, and/or latitudinal motions of the tail. The average $|B_X|$ of the high latitude tail over the entire observable cross section is 10.64γ and its distribution is slightly non-uniform, similar to that of B.

Figure 5 shows the distribution of the tail field deviation along the east-west direction ($Y_{SM}$ axis). The lower value in each element is the observed $B_Y$ which is the deviation with respect to the Sun-Earth line ($X_{SM}$ axis) and the upper $B_Y$ includes the tail axis aberration correction of ± 0.6γ in ±Z region corresponding to an average field of 11γ and an aberration angle of about $3.1^{\circ}$. It is interesting to note that after the correction, $B_Y$ in the northern-dawn quadrant changed from negative to positive and this corrected spatial distribution of $B_Y$ (the east-west longitudinal deviation of the tail field with respect to the aberrated tail axis) is qualitatively consistent with the configuration of field lines for a diverging magnetotail, namely negative $B_Y$'s in the northern-dusk and southern-dawn quadrants and positive ones in the northern-dawn and southern-dusk quadrants.

The distribution of $|B_Y|$, which represents the magnitude of

the longitudinal deviation from the Sun-Earth line, is not uniform. In the dusk part of the tail, $|B_Y|$ gradually decreases from about $2\gamma$ near the flank to about $1\gamma$ at the center of the tail; however, in the dawn side it maintains a constant value of about 0.8 or $0.9\gamma$. In latitude, $|B_Y|$ has a greater value of $1.8\gamma$ at $Z_N \approx \pm 18\ R_E$ than that near $Z_N \approx 0$ of $1.2\gamma$.

In order to see quantitatively the longitudinal divergence of the tail field along the $Y_{SM}$ direction, the distribution of the longitudinal angle $|\varphi|$ is shown in Figure 6, in which $|\varphi|$ is the absolute angular deviation of $\varphi$ with respect to $0^\circ$ and $180^\circ$ axis. Along the Y-direction, $|\varphi|$ is gradually decreasing from $12^\circ$ at $Y_{SM} \approx 24\ R_E$ to $4.5^\circ$ at $Y_{SM} \approx -3\ R_E$, indicating an approximate divergence rate of about $1^\circ$ per $3\ R_E$ or higher from the tail center to the evening side of the magnetopause. However, in the morning side of the tail, $|\varphi|$ has nearly a constant value of $4.5^\circ$, indicating that the magnetic field lines may not deviate as much as those in the evening side and that the divergence is about $7.6^\circ$ with respect to the aberrated axis of the magnetotail. Along the Z-direction, $|\varphi|$ does not distribute symmetrically in $Z_N$, with the smallest $|\varphi|$ near $Z_N \approx -5\ R_E$, and it is always larger in the northern tail than in the southern. This tendency indicates a greater degree of longitudinal divergence in the northern part of the tail than in the southern for the data sampled during winter and summer, respectively.

The latitudinal (north-south) divergence of the magnetic field lines in the meridian planes is illustrated by the distribution of $B_Z$ component and latitude angle $\theta$ in Figure 7. The profile of averaged

$\theta$ over different $Z_N$ values shows that $\theta$ is smoothly decreasing from about $+3^o$ at $Z_N \approx 0\ R_E$ to about $-3^o$ at $Z_N \approx \pm 15\ R_E$; and the tail field line divergence exists, on the average, in regions above $Z_N \approx 6\ R_E$ and below $Z_N \approx -12\ R_E$, indicated by negative values of $B_Z$'s and $\theta$'s. Along the $Y_{SM}$ axis, $\theta$ is $-1.5^o$ near the dusk flank and gradually increases to $+3^o$ near the dawn flank, revealing a non-uniform latitudinal divergence along Y, even though the $|\theta|$ has an almost constant value of about $4^o$ over the whole tail cross section. In the magnetotail, the convergence or divergence of magnetic field lines with respect to the magnetotail equatorial plane is represented by a positive or negative $\theta$ angle, respectively; thus, the latitudinal divergence can also be determined from the distribution of the occurrence frequency for negative $\theta$ angles (Figure 8). Between $Z_N = +3\ R_E$ and $-3\ R_E$, only 20% of the observed field lines diverges, gradually increasing to 70% at $|Z_N| \approx 18\ R_E$. Thus, the field lines are essentially converging near the magnetotail equatorial plane, but they have a greater tendency of divergence at a greater distance from the neutral sheet. Along the Y-direction, about 60% of the $\theta$ angles are negative near the dusk flank and slowly decreases to only 30% near the dawn flank, corresponding to a greater degree of the latitudinal (north-south) divergence in the dusk side. Therefore, the field line divergence, in both longitudinal (east-west) and latitudinal (north-south) directions, is larger in the evening part of the tail than in the morning side.

B. Temporal variations

There are two major causes for the magnetotail configurations

to change: the solar wind (and its related geomagnetic activity) and the season of a year. Since the comprehensive solar wind conditions are not available during these 34 tail crossings, the magnetic field measurements in the high latitude tail lobes are divided either into the low $K_p$ ($K_p \leq 1+$) and high $K_p$ ($K_p \geq 2-$) subgroups according to the geomagnetic activity, or three seasonal subgroups, equinox, summer and winter. The ratio of the number of data points between low $K_p$ and high $K_p$ groups is about 3 to 2 and seasonally is about 2:1:1.

### 1. $K_p$ dependence

The most obvious tail field variations associated with the $K_p$ index is the field magnitude. The magnitude (B) (see Figure 2) and the absolute values of two field components ($|B_X|$, $|B_Z|$) are about 87% for low $K_p$ ($\leq 1+$) values and 120% for high $K_p$ ($\geq 2-$) values with respect for all $K_p$ values. Angularly, the average of the latitudinal angle ($\theta$) over the entire cross section is quite different ($0.4^o$ for low $K_p$ and $1.2^o$ for high $K_p$), but $|\theta|$ is similar ($\sim 4.3^o$) for different $K_p$ values (see Table 1). On the other hand, from the distributions of the $|\omega|$ along the Y-direction (see Figure 6) it is found that there are no significant differences between high $K_p$ and low $K_p$ groups, indicating statistically no change in the longitudinal (east-west) angle during a moderate or an intense geomagnetic activity.

### 2. Seasonal variations

In studying seasonal variations of the magnetotail at the lunar orbit, it is important to note that due to the intrinsic limitation of the

lunar orbit, the Explorer 35 observations are confined to the southern part of the tail centered at $Z_N \approx -8\ R_E$ in summer months, and to the northern tail centered at $Z_N \approx +8\ R_E$ in winter months. In summer, due to the presence of more high $K_p$ observations, the average field magnitude is higher (at 12.5 $\gamma$) than the all year average of 10.8$\gamma$, and so is the absolute value of its components ($B_X$, $B_Y$, $B_Z$). Directionally, $|\theta|$ does not show any significant seasonal variations; however, averaged $|\varphi|$ is smaller in summer ($\sim 6.3^\circ$) than in winter ($\sim 7.9^\circ$) and the profiles of $|\varphi|$ distribution (Figure 9) are different between two solstices, namely in the evening side $|\varphi|$ values are about $3^\circ$ or greater in winter than in summer. This apparent seasonal difference is actually due to a spatial variation in $|\varphi|$ distribution between northern and southern tail lobes, since the data from equinoctial months also show that $|\varphi|$ is larger in the northern-dusk quadrant than in the southern-dusk. During equinox, Explorer 35 measurements are centered at $Z_N \approx 0\ R_E$ and the observed distributions of high latitude tail lobe field line configuration are essentially the same as those without seasonal and $K_p$ separation.

The most interesting feature of the high latitude field lines is their divergence with respect to the tail axis. It is clearly illustrated by the longitudinal $|\varphi|$ angles observed during summer months in the southern tail lobe. Figure 10 shows the average $\varphi$ value at each 3 $R_E$ wide $Y_{SM}$ interval, together with the $\varphi$ angle longitudinal distribution detected between $-9\ R_E \leq Z_N \leq -6\ R_E$. The divergence angle between the field lines near the dusk and dawn flanks of the tail is about $18^\circ$ and the approximate divergence rate is

larger ($\sim 0.7^{\circ}$ per $R_E$) in the dusk side than that ($\sim 0.3^{\circ}$ per $R_E$) in the dawn side. This gives a model configuration of the field lines in the high latitude tail region along the azimuthal direction from $\varphi = 168^{\circ}$ at $Y_{SM} \approx 18\ R_E$, gradually increasing to $180^{\circ}$ at $Y_{SM} = 0$ and to $184^{\circ}$ at $Y_{SM} \approx -18\ R_E$ in the southern tail.

The characteristics of the magnetic field lines in the high latitude tail region near 60 $R_E$ observed by Explorer 35 are summarized in Table 1. It clearly indicates that the magnetic field lines diverge in both longitudinal ($\varphi$) and latitudinal ($\theta$) directions. The degree of divergence is greater in the evening side of the magnetotail than that in the morning side, and greater in the northern tail than in the southern. The divergence angle between field lines near the dusk flank and those near the dawn flank is about $18^{\circ}$, but the divergence angle along the Z-direction is only about $8^{\circ}$; thus, the rate of divergence is larger along the east-west direction than along the north-south direction. It is also found that the configuration of the high latitude magnetic field lines is not affected by geomagnetic activity (or the solar wind pressure variations).

## IV. Plasma sheet

There are about 1500 hours of the Explorer 35 magnetic field measurements which are taken inside the plasma sheet at the lunar orbit. Data sampling points are evenly distributed either in the northern versus the southern or the dusk versus the dawn sides of the magnetotail as shown in Figure 1. The magnetic field in the plasma sheet is characterized by the diamagnetic effect of plasma, the angular

deviations away from the Sun-Earth alignment and above all, the frequent large fluctuations. The scatter of data with respect to the average (i. e. arithmetic mean) of each field parameter is rather large as illustrated by the percentage distribution of three magnetic field components in Figure 11. It is found that the average of field magnitude (B) over the entire tail cross section of the plasma sheet is 7.81$\gamma$ which is about 30% less than that of the high latitude tail lobe due to the plasma diamagnetism. The average of the latitudinal angle ($\theta$) is 10.57$^{o}$, indicating a large angular deviation from the solar magnetospheric equatorial plane, and the standard deviation with respect to the average is 19.29$^{o}$, which is almost four times greater than that observed in the high latitude tail, corresponding to a large angular fluctuation inside the plasma sheet.

## A. Spatial distribution

The spatial distribution of the magnetic field components observed within the plasma sheet is not usually systematic as that observed outside the plasma sheet is; for example, in the distribution of the field magnitude (B), there is no indication for small B's located near the equatorial plane or any other trends in the distribution. Similarly, the field strength in the X-direction, $|B_X|$, is randomly distributed with an average of 6.73$\gamma$. However, except for a few anomalous points, the direction of the averaged magnetic field X-component, $B_X$, within the plasma sheet is sunward (positive) in the northern tail and anti-sunward (negative) in the southern tail as shown in Figure 12; and the mean value of $B_X$ over the entire cross

section is almost zero (0.04$\gamma$), indicating balanced observations above and below the neutral sheet. Values of $B_X$ (upper one in each block) and $|B_X|$ (lower one) observed inside the plasma sheet are different in each areal element; this is not the case in the high latitude tail. This means that the neutral sheet, separating two oppositely directed field line regions, is not stably located adjacent to the calculated $Z_N = 0$ location and is frequently observed even at a distance of more than 10 $R_E$ away from the expected location.

The $B_Y$ values are almost all negative over the entire cross section with an average value of -0.69$\gamma$, corresponding to an additional eastward (dusk to dawn) tail field component; its spatial distribution does not show any indication of ordered field lines diverging, even after the tail axis aberration correction. The average of $|B_Y|$ in the plasma sheet is 2.01$\gamma$, which is about 50% greater than that observed in the high latitude tail.

The north-south component of the magnetic field ($B_Z$) has an average value of 1.09$\gamma$ inside the plasma sheet and its spatial distribution reveals that large $B_Z$ values are detected near the tail equatorial plane and also in the morning side. The magnitude of the field across the neutral sheet is best represented by $|B_Z|$ which has an average value of 1.91$\gamma$ over the entire plasma sheet, and Figure 13 illustrates its distribution along Y and Z-directions. Longitudinally, the $|B_Z|$ is approximately constant with 1.6$\gamma$ between $Y_{SM} = 15\ R_E$ and -12 $R_E$ and higher near both flanks; latitudinally, it is about constant between $Z_N = \pm\ 12\ R_E$ and slightly enhanced beyond this region.

Angularwise, it is found that the average longitudinal deviation

$|\varphi|$ of field lines from the Sun-Earth direction is 19.65°, which is almost three times larger than that in the high latitude tail lobe. The spatial distribution of $|\varphi|$ has a "w" shaped profile along the Z-direction; it is about 25° at $|Z_N| \geq 15\ R_E$ and ~18° at $|Z_N| \approx 9\ R_E$ with a value of ~20° near $Z_N \approx 0\ R_E$. Along $Y_{SM}$ axis, it is about 22° near the dusk flank and gradually decreases to ~16° near $Y_{SM} \approx -6\ R_E$ but then increases to ~25° at $Y_{SM} \approx -20\ R_E$ (Figure 14).

The observed latitude angle $\theta$ inside the plasma sheet is mostly (73.3%) positive with an average value of 10.6°, indicating that the magnetotail neutral point is generally located beyond the lunar orbit. The spatial distribution of $\theta$ shows that latitudinally large $\theta$ values are generally distributed near the tail equatorial plane ($|Z_N| \leq 6\ R_E$) and longitudinally small $\theta$ values are in the center of the tail. Figure 15 illustrates the longitudinal and latitudinal variations of $|\theta|$, which indicates the true north-south angular deviation of the field lines, with a value of ~12.5° at the tail center ($Y_{SM} \approx -3\ R_E$), increasing toward both flanks along Y and approximately with a constant value which is very close to the average of ~19.65° along Z axis.

B. Temporal variations

Among the 8791 10-minute magnetic field data obtained in the plasma sheet, 63% of them are observed during periods of low geomagnetic activity ($K_p \leq 1+$) and the rest in active periods ($K_p \geq 2-$). It is found that the spatial distributions of all magnetic field para-

meters, as well as their absolute values, do not change significantly with $K_p$ values. Magnitudewise, however, the field strength B shows the largest variation, increasing from 6.77γ at low $K_p$ to 9.56γ at high $K_p$ (see Figure 11); and consequently $|B_X|$, $B_Y$, $|B_Y|$, $B_Z$ and $|B_Z|$ are also larger during high $K_p$ than low $K_p$. The change of the field line configuration (i. e. direction) can be determined from the observations of the angles, θ and φ. The absolute longitude angle $|\varphi|$ does not have a significant change (19.88° at low $K_p$ and 19.27° at high $K_p$) with geomagnetic activity. However, the average latitude angle θ is 11.64° at low $K_p$ and 8.79° at high $K_p$; the average of absolute value of latitudinal angle $|\theta|$ is 17.02° during quiet times and 15.80° during geomagnetically active periods. The occurrence frequency of a positive θ angle decreases slightly from 74.6% at low $K_p$ to 71.2% at high $K_p$ (Figure 16). Thus, the field line configuration variation associated with geomagnetic activity is found only in the latitude angle θ and its absolute value $|\theta|$, namely they are larger during geomagnetically quiet times than active periods.

Seasonally, the field magnitude is about 10% greater in summer than in winter, due to a coincidental uneven $K_p$ distribution; thus, the three field components are also higher in summer than in winter. Angularly, θ, as well as $|\theta|$, are larger (12.73° and 16.16°, respectively) in summer than in winter (8.5° and 15.32°); however, $|\varphi|$ is about the same for two solstice seasons. The ratio of positive to negative θ angles is quite different, from 83.7/16.1 in summer months to 66.2/33.6 in winter months.

Table 2 summarizes the magnetic field line characteristics

observed inside the plasma sheet. It is clear that inside the plasma sheet the angular deviation of field lines from the Sun-Earth direction is greater by a factor of three or four than that in the high latitude tail lobes, and the spatial distribution of the magnetic field parameters do not show any indication of a systematic diverging field line structure. The average configuration of field lines does not vary drastically during different geomagnetic activities, but the magnitude of north-south directional deviation, indicated by $|\theta|$, is greater during quiet periods than active periods.

## V. Conclusion and discussion

The magnetotail field line configurations near 60 $R_E$ are studied extensively in the high latitude tail regime and inside the plasma sheet, by using the magnetic field measurements (34 magnetotail crossings) from the lunar orbiting Explorer 35 satellite.

In the high latitude tail lobes (i. e. outside the plasma sheet), the magnetic field lines are systematically and orderly distributed, with a clear indication of their divergence along the magnetotail. From the distribution of the $\varphi$ angle, it is found that the field lines gradually diverge longitudinally away from the aberrated tail axis and the diverging angle is about $18^o$ between two flanks with a higher degree of divergence in the dusk side than in the dawn side and in the northern tail than in the southern tail. This $18^o$ longitudinal divergence is in good agreement with the earlier result of Behannon (1970).

The divergence along the latitudinal direction is also observed,

as indicated by negative $\theta$ angles. It is located mainly above $Z_N \approx 3\ R_E$ and below $Z_N \approx -9\ R_E$, but field lines with positive $\theta$ angles (corresponding to a convergent configuration) are observed between them. The intrinsic limitation of the lunar orbit confines the Explorer 35 observations to only about $\pm 6\ R_E$ in the Z-direction near the midnight meridian, thus, the exact latitudinal divergence at large $|Z|$ values near $Y_{SM} \approx 0$ cannot be determined. Since $|\theta|$ does not depend on $Y_{SM}$ and since $|\theta|$ deviation is about $5^\circ$ at $|Z_N| \approx 18\ R_E$ (together with the fact that $|Y_{SM}|$ values are large), it is estimated that the north-south tail divergence is about $10^\circ$.

The fact that the field line divergence angle observed along the Y-direction is about twice as that along the Z-direction reveals that the shape of the tail cross section should be elliptical with the semi-major axis oriented along the solar magnetospheric Y-axis. This result is quite differnt from the previous report by Behannon (1970) of the semi-major axis along the Z-direction; his result was based on the Explorer 33 magnetopause crossings only at large -Z values.

Inside the plasma sheet, the field lines are angularly deviated along the dusk-to-dawn direction and converging toward the neutral sheet; and the systematic tail divergence does not exist. The longitudinal (east-west) deviation angle $|\varphi|$ is about $20^\circ$, corresponding to $|B_Y|$ of about $2\gamma$ and the latitudinal (north-south) deviation $|\theta|$ is also very large of about $17^\circ$. It is also found that $|B_Z|$ is about $1.9\gamma$ without $K_p$ selection and $1.69\gamma$ for $K_p \leq 1+$ which are much larger than the $|B_Z|$ reported earlier with a value of $\sim 0.65\gamma$ (Behannon, 1970).

This large increase is due to a clear separation of measurements inside the plasma sheet and those outside of it. This newly observed large $|B_Z|$ indicates that the gradient of the tail field magnitude along the X-direction should be greater than the previously estimated one ($B \propto |X|^{-0.3}$).

The $K_p$ dependence of the field line configuration is not obvious in the high latitude tail lobe, but inside the plasma sheet the latitudinal deviation $|\theta|$ shows a decrease from $17.24^\circ$ at low $K_p$ to $15.80^\circ$ at high $K_p$. Although the standard deviations with respect to these mean values are large (of about $14.4^\circ$) due to the fluctuating nature of the field in the plasma sheet, this decrease is still considered meaningful, since it is consistently detected at different Y and Z intervals. It has long been though that the magnetotail field lines become more dipole-like during the expansion phase of magnetospheric substorms (see for example, McPherron, 1972; Nishida and Nagayama, 1973); thus, the magnetic field line configuration is expected to have a larger latitudinal deviation $|\theta|$ during high $K_p$ than during low $K_p$. However, the results in this analysis show that in the high latitude tail lobe near 60 $R_E$, $|\theta|$ does not change significantly with $K_p$ values; this is in good agreement with an earlier observation in the distant tail region (Meng et al., 1971). Further, inside the plasma sheet, $|\theta|$ is about $1.5^\circ$ larger at low $K_p$ than at high $K_p$. This tendency is opposite to what one would expect from the proposed change from a tail-like to a dipole-like field line configuration during substorms, unless this substorm variation is statistically masked by other temporal variations of the magnetotail configuration.

It is also interesting to examine the formation of a new neutral

line (or lines) inside the cislunar distance during substorms, as reported by Nishida and Nagayama (1973) and Nishida and Hones (1974). In Section IV it was shown that the percentage of the occurrence of negative $\theta$ angle inside the plasma sheet is 28.6% at high $K_p$, compared to 25.3% during low $K_p$. This observed 3.3% increase in the occurrence frequency of negative $\theta$ angles seems to be too small to support the speculation that the new neutral line formation inside the lunar orbit is a common feature during substorms. Furthermore, the $K_p$ dependence on the occurrence frequency of negative $\theta$ values is practically nil in the center of the plasma sheet (i.e. small $|Z_N|$ and $|Y_{SM}|$) (see Figure 16).

Inside the plasma sheet, the ratio of the number of positive-negative $\theta$ angles shows two puzzling features. First of all the occurrence frequency of negative $\theta$ angles is 16.1% in summer and becomes doubled (33.6%) in winter. This drastic increase in the winter season could not be explained as due to the seasonal change of the dipole axis orientation and of the plasma sheet location with respect to the lunar orbiting plane, but it may be attributed to the differences in the lunar orbiting distances across the tail during these two solstice seasons. In summer, the lunar orbit is at about $X = -57\ R_E$ in the tail, but it is further downstream in the tail at $X \approx -63\ R_E$ (i.e. closer to the average location of the neutral line) in winter. The other puzzling observation is the very low occurrence frequency of negative $\theta$ angles located in the region $-18\ R_E \leqslant Y_{SM} \leqslant -6\ R_E$, shown in Figure 16. At this $Y_{SM}$ interval, very high energetic electron fluxes have been observed by Vela at $X \approx -18\ R_E$ and by Explorer 35 satellites at lunar orbit (Montgomery, 1968; Meng, 1971). However, whether this low

occurrence frequency of negative $\theta$ values has any relationship with the large energetic electron fluxes requires further study.

## ACKNOWLEDGMENTS

This work was supported by the Atmospheric Sciences Section of the National Science Foundation under Grant GA-26582A1 and the National Aeronautics and Space Administration under Grant NGL-05-003-017. The authors are grateful to S. I. Akasofu for his helpful suggestions during this investigation and J. D. Mihalov, D. S. Colburn and C. P. Sonett for providing the Ames magnetometer data.

# REFERENCES

Behannon, K. W., Mapping of the earth's bow shock and magnetic tail by Explorer 33, J. Geophys. Res., 73, 907, 1968.

Behannon, K. W., Geometry of the geomagnetic tail, J. Geophys. Res., 75, 743, 1970.

Behannon, K. W., and N. F. Ness, Satellite studies of the earth's magnetic tail, in Physics of the Magnetosphere, edited by R. L. Carovillano, J. F. McClay, and H. R. Radoski, p. 409, D. Reidel Publishing Company, Dordrecht, Holland, 1969.

Hruska, A., and J. Hruskova, Transverse structure of the earth's magnetotail and fluctuation of the tail magnetic field, J. Geophys. Res., 75, 2449, 1970.

McPherron, R. L., Substorm related changes in the geomagnetic tail: The growth phase, Planet. Space Sci., 20, 1521, 1972.

Meng, C. I., Energetic electrons in the magnetotail at 60 $R_E$, J. Geophys. Res., 76, 862, 1971.

Meng, C. I., and J. D. Mihalov, Average plasma sheet configuration near 60 earth radii, J. Geophys. Res., 77, 1739, 1972a.

Meng, C. I., and J. D. Mihalov, On the diamagentic effect of the plasma sheet near 60 $R_E$, J. Geophys. Res., 77, 4661, 1972b.

Mihalov, J. D., D. S. Colburn, R. G. Currie, and C. P. Sonett, Configuration and reconnection of the geomagnetic tail, J. Geophys. Res., 73, 943, 1968.

Mihalov, J. D., C. P. Sonett, and D. S. Colburn, Reconnection and noise in the geomagnetic tail, Cosmic Electrodyn., 1, 178, 1970.

Montgomery, M. D., Observations of electrons in the earth's magnetotail by Vela 2, J. Geophys. Res., 73, 871, 1968.

Murayama, T., Spatial distribution of energetic electrons in the geomagnetic tail, J. Geophys. Res., 71, 5547, 1966.

Ness, N. F., The earth's magnetic tail, J. Geophys. Res., 70, 2989, 1965.

Ness, N. F., K. W. Behannon, C. S. Scearce, and S. C. Cantarano, Early results from the magnetic field experiment on lunar Explorer 35, J. Geophys. Res., 72, 5769, 1967.

Nishida, A., and N. Nagayama, Synoptic survey for the neutral line in the magnetotail during the substorm expansion phase, J. Geophys. Res., 78, 3782, 1973.

Nishida, A., and E. W. Hones, Jr., Association of plasma sheet thinning with neutral line formation in the magnetotail, J. Geophys. Res., 79, 535, 1974.

Russell, C. T., and K. I. Brody, Some remarks on the position and shape of the neutral sheet, J. Geophys. Res., 72, 6104, 1967.

Sonett, C. P., D. S. Colburn, R. G. Currie, and J. D. Mihalov, The geomagnetic tail: Topology, reconnection and interaction with the moon, in Phyiscs of the Magnetosphere, edited by R. L. Carovillano, J. F. McClay, and H. R. Radoski, p. 461, D. Reidel Publishing Company, Dordrecht, Holland, 1969.

Table 1 Field line components observed in the high latitude tail lobe

| Parameter | Magnitude | | | | | | Distribution[5] |
|---|---|---|---|---|---|---|---|
| | Average | $K_p \leq 1^+$ | $K_p \geq 2^-$ | Equinox | Summer[2] | Winter[3] | |
| B | 10.8γ | 9.4γ | 13.2γ | 10.2γ | 12.5γ[1] | 10.4γ | Greater for small $\|Y\|$<br>Greater in the southern tail[1] |
| $B_X$ | 0.15γ | 0.75γ | -3.57γ | -1.4γ | -11.6γ[2] | 9.5γ[3] | Greater for small $\|Y\|$<br>$+B_X$ in +Z; $-B_X$ in -Z region |
| $\|B_X\|$ | 10.6γ | 9.3γ | 13.0γ | 10.0γ | 12.4γ[1] | 10.2γ | same as B |
| $B_Y$ | -0.30γ | -0.44γ | -0.06γ | -0.24γ | +0.59γ | -1.19γ | $-B_Y$ in (+Y, +Z) and (-Y, -Z) regions[4]<br>$+B_Y$ in (-Y, +Z) and (+Y, -Z) regions[4] |
| $\|B_Y\|$ | 1.3γ | 1.2γ | 1.5γ | 1.2γ | 1.4γ | 1.4γ | Greater near dusk flank, decreasing toward $Y \approx 0$, constant in -Y; greater at large $\|Z_N\|$ |
| $B_Z$ | 0.15γ | 0.06γ | 0.3γ | 0.1γ | 0.4γ | -0.02γ | $B_Z < 0$ at +Y, $B_Z > 0$ at -Y<br>$B_Z < 0$ at large $\|Z_N\|$, $B_Z > 0$ at $-10 \leq Z_N \leq 5\ R_E$ |
| $\|B_Z\|$ | 0.8γ | 0.7γ | 1.0γ | 0.8γ | 0.9γ | 0.7γ | uniform over Y's and Z's. |
| φ | no meaning | | | | 181° | | <180° in (+Y, -Z) region, >180° in (-Y, -Z)<br>>0° in (-Y, +Z) region, <360° in (+Y, +Z) |
| $\|\phi\|$ | 7.1° | 7.3° | 6.6° | 7.0° | 6.3° | 7.9° | Greater near dusk flank, decreasing toward $Y \approx 0$, constant in -Y. Greater at large $\|Z_N\|$, smaller at $Z_N \approx -5\ R_E$. |
| θ | 0.7°<br>+θ 55%<br>-θ 44% | 0.4°<br>+θ 53%<br>-θ 46% | 1.2°<br>+θ 59%<br>-θ 40% | 0.5°<br>+θ 54%<br>-θ 45% | 2.1°<br>+θ 67%<br>-θ 32% | -0.2°<br>+θ 47%<br>-θ 52% | Negative in +Y, positive in -Y,<br>Negative at large $\|Z_N\|$, positive at $-12 \leq Z_N \leq +6\ R_E$ |
| $\|\theta\|$ | 4.29° | 4.31° | 4.25° | 4.5° | 4.1° | 4.1° | Approximately uniform over Y's and Z's. |

Notes:
[1] Likely due to $K_p$ effect, not true spatial or seasonal variation
[2] Observations limited in the southern tail
[3] Winter observations limited in the northern tail
[4] After magnetotail axis aberration correction
[5] No change with $K_p$ and season

Table 2 Field line components observed inside the plasma sheet

| Parameter | Magnitude | | | | | | Distribution[5] |
|---|---|---|---|---|---|---|---|
| | Average | $K_p \leq 1^+$ | $K_p \geq 2^-$ | Equinox | Summer[2] | Winter[3] | |
| B | 7.81γ | 6.77γ | 9.56γ | 7.39γ | 8.19γ[1] | 8.08γ | approximately uniform |
| $B_X$ | 0.04γ | 0.59γ | -0.89γ | -0.35γ | -5.62γ[2] | 4.43γ[3] | $+B_X$ in the +Z region<br>$-B_X$ in the -Z region |
| $\lvert B_X \rvert$ | 6.73γ | 5.83γ | 8.25γ | 6.25γ | 7.82γ[1] | 7.13γ | approximately uniform |
| $B_Y$ | -0.69γ | -0.70γ | -0.68γ | -0.56γ | -0.45γ | -1.12γ | $-B_Y$ over the whole cross section |
| $\lvert B_Y \rvert$ | 2.01γ | 1.79γ | 2.38γ | 1.94γ | 2.23γ | 2.02γ | larger at high Z ($\lvert Z \rvert \geq 12\ R_E$)<br>larger near both flanks |
| $B_Z$ | 1.09γ | 1.07γ | 1.13γ | 1.15γ | 1.54γ | 0.69γ | larger near the $Z_N \approx 0$ plane<br>larger in the dusk side of the tail |
| $\lvert B_Z \rvert$ | 1.91γ | 1.69γ | 2.28γ | 1.94γ | 2.08γ | 1.75γ | larger in the morning side<br>about constant between $Z_N \approx \pm 12\ R_E$<br>slightly higher beyond $\lvert Z_N \rvert \leq 12\ R_E$ |
| $\Phi$ | no meaning | | | | 188.9°[6] | | |
| $\lvert \Phi \rvert$ | 19.7° | 19.88° | 19.27° | 20.70° | 18.20° | 18.26° | larger near both flanks<br>larger at high Z ($\lvert Z \rvert \geq 12\ R_E$) |
| θ | 10.57°<br>+θ 73.3%<br>-θ 26.5% | 11.64°<br>+θ 74.6%<br>-θ 25.3% | 8.79°<br>+θ 71.2%<br>-θ 28.6% | 10.91°<br>+θ 73.6%<br>-θ 26.3% | 12.73°<br>+θ 83.7%<br>-θ 16.1% | 8.50°<br>+θ 66.2%<br>-θ 33.6% | larger near the equatorial plane ($\lvert Z_N \rvert \leq 6\ R_E$)<br>larger near both flanks |
| $\lvert \theta \rvert$ | 16.56° | 17.02° | 15.80° | 17.24° | 16.16° | 15.32° | approximately constant along $Z_N$<br>larger near both flanks |

Notes: see Table 1

Figure Captions

Figure 1. Spatial distribution of 10-min average magnetic field data. The upper and lower values in each 3 $R_E \times 3\ R_E$ square are the number of field measurements obtained in the high latitude tail lobe (i. e. outside the plasma sheet) and inside the plasma sheet, respectively.

Figure 2. Distribution of the magnetic field magnitude observed in the high latitude tail during different geomagnetic activities. The arrow indicates the average value.

Figure 3. Longitudinal and latitudinal distributions of the high latitude field magnitude.

Figure 4. Spatial distribution of the high latitude field X-component. The upper value is the average of the $B_X$ in that areal element and the lower value is the average of the $|B_X|$.

Figure 5. Spatial distribution of the high latitude field Y-component. Lower values are the averages of the observed $B_Y$. Upper values include the magnetotail axis aberration correction of $\pm$ 0.6$\gamma$ in $\pm$ Z; thus, they represent a distribution of $B_Y$ with respect to the aberrated magnetotail axis.

Figure 6. Longitudinal and latitudinal distributions of azimuthal deviation of the high latitude tail field lines with respect to the sun-earth alignment. $|\varphi|$ refers to the absolute angular deviation of $\varphi$ from $0^\circ$ or $180^\circ$ axis. Please note that a gradually increasing $|\varphi|$ in the dusk side (+Y's) represents a "fan shaped" field line divergence away from the center of the tail.

Figure 7. Longitudinal and latitudinal distributions of the latitude angle ($\theta$) and $B_Z$-component. A positive and negative $\theta$ (or $B_Z$) represents the north-south convergence and divergence, respectively, of the field lines with respect to the solar magnetospheric equatorial plane (or the neutral sheet).

Figure 8. Distribution of the occurrence frequency for negative $\theta$ angles (i. e. latitudinal diverging field lines).

Figure 9. The longitudinal distribution of the absolute azimuthal angular deviations away from the sun-earth line observed in the high latitude tail during summer and winter months. The deviation in the dusk side is larger in winter than in summer. This is owing to the asymmetric field line divergence in the northern and southern tail, since the summer and winter observations were limited in the southern and northern magnetotail, respectively.

Figure 10. The longitudinal distribution of the high latitude field line azimuthal angle ($\varphi$) observed in the summer months (i. e. only in the southern tail). The fan shaped longitudinal divergence of field lines away from the center of the tail is rather obvious and it is clear that the rate of divergence is greater in the dusk side than in the dawn side of the tail.

Figure 11. Distributions of the magnetic field magnitude, $B_Y$ and $B_Z$ with respect to thier averaged values, observed inside the plasma sheet. The scatter of B is larger inside the plasma sheet than in the high latitude tail as shown in Figure 2.

Figure 12. Spatial distributions of the plasma sheet magnetic field component $B_X$ and $|B_X|$. Longitudinal and latitudinal distributions of $|B_X|$ are also shown on the top and right hand side of the spatial distribution.

Figure 13. Longitudinal and latitudinal distributions of the plasma sheet field intensity across the neutral sheet ($|B_Z|$) during different geomagnetic conditions.

Figure 14. Similar to Figure 13, but for the absolute azimuthal angular deviations from sun-earth alignment observed inside the plasma sheet.

Figure 15. Similar to Figure 13, but for the absolute latitudinal (north-south) angular deviation from the solar magnetospheric equatorial plane observed inside the plasma sheet.

Figure 16. Similar to Figure 13, but for the occurrence frequency of negative $\theta$ angles inside the plasma sheet.

# DISTRIBUTIONS OF MAGNETIC OBSERVATION

Upper No.: High Latitude Lobe  Lower No.: Plasma Sheet

| $Z_N$ | 24 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 | -3 | -6 | -9 | -12 | -15 | -18 | -21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 8<br>40 | 32<br>21 | 33<br>20 | 25<br>14 | | | | | | | | | 9<br>9 | 8<br>42 | 13<br>31 | |
| 15 | 36<br>44 | 91<br>86 | 32<br>19 | 55<br>27 | 20<br>5 | | | | | | 10<br>15 | 41<br>36 | 48<br>54 | 12<br>40 | 9<br>48 | |
| 12 | 67<br>79 | 34<br>73 | 22<br>46 | 60<br>59 | 100<br>43 | 123<br>56 | 35<br>47 | 18<br>0 | 120<br>11 | 76<br>15 | 79<br>14 | 41<br>37 | 32<br>64 | 39<br>41 | 45<br>115 | |
| 9 | 19<br>61 | 48<br>37 | 48<br>85 | 125<br>126 | 103<br>114 | 72<br>30 | 88<br>67 | 232<br>106 | 92<br>86 | 103<br>46 | 87<br>100 | 110<br>78 | 27<br>23 | 29<br>64 | 53<br>90 | |
| 6 | 12<br>48 | 24<br>74 | 43<br>91 | 29<br>39 | 36<br>74 | 76<br>119 | 51<br>101 | 75<br>96 | 55<br>74 | 45<br>131 | 40<br>102 | 54<br>119 | 29<br>114 | 35<br>115 | 15<br>16 | |
| 3 | 10<br>25 | 38<br>106 | 10<br>69 | 23<br>81 | 17<br>55 | 77<br>68 | 23<br>63 | 29<br>100 | 25<br>72 | 34<br>151 | 21<br>108 | 8<br>50 | 28<br>113 | 14<br>80 | 2<br>38 | |
| 0 | 3<br>54 | 18<br>79 | 14<br>21 | 14<br>58 | 12<br>83 | 26<br>87 | 57<br>76 | 52<br>76 | 14<br>79 | 42<br>121 | 15<br>60 | 30<br>75 | 19<br>82 | 12<br>97 | 1<br>54 | |
| -3 | 16<br>28 | 29<br>56 | 38<br>39 | 61<br>45 | 52<br>34 | 90<br>113 | 68<br>23 | 82<br>93 | 103<br>96 | 73<br>100 | 98<br>77 | 27<br>143 | 28<br>54 | 31<br>69 | 3<br>58 | |
| -6 | 15<br>50 | 32<br>84 | 51<br>81 | 78<br>56 | 113<br>51 | 78<br>55 | 73<br>8 | 153<br>40 | 97<br>34 | 137<br>37 | 90<br>60 | 52<br>89 | 97<br>84 | 30<br>105 | 27<br>61 | |
| -9 | 26<br>67 | 96<br>39 | 63<br>35 | 82<br>21 | 70<br>16 | 99<br>14 | 79<br>18 | 61<br>11 | 15<br>14 | 38<br>17 | 55<br>7 | 58<br>12 | 29<br>46 | 62<br>60 | 22<br>24 | |
| -12 | 13<br>8 | 92<br>47 | 66<br>16 | 68<br>6 | 18<br>2 | 5<br>0 | | | | | | 8<br>5 | 25<br>8 | 49<br>49 | 4<br>27 | |
| -15 | 12<br>24 | 55<br>41 | 12<br>13 | 21<br>2 | | | | | | | | | | 4<br>54 | 0<br>26 | |
| -18 | | | | | | | | | | | | | | | | |

$Y_{SM}$

Figure 1

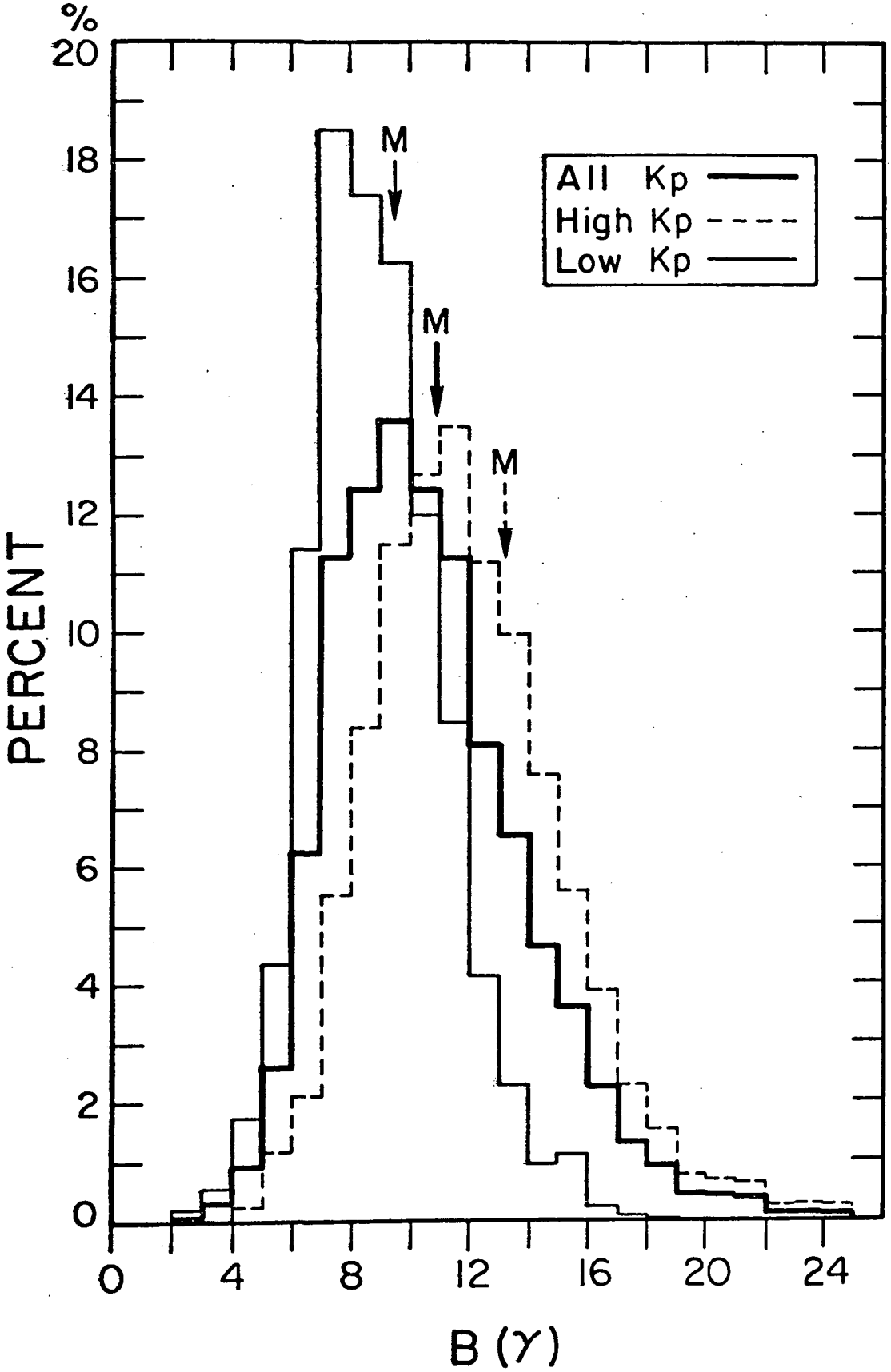

%
20
18
16
14
12
10
8
6
4
2
0
PERCENT
M
M
M
All Kp
High Kp
Low Kp
0
4
8
12
16
20
24
B (γ)

Figure 2

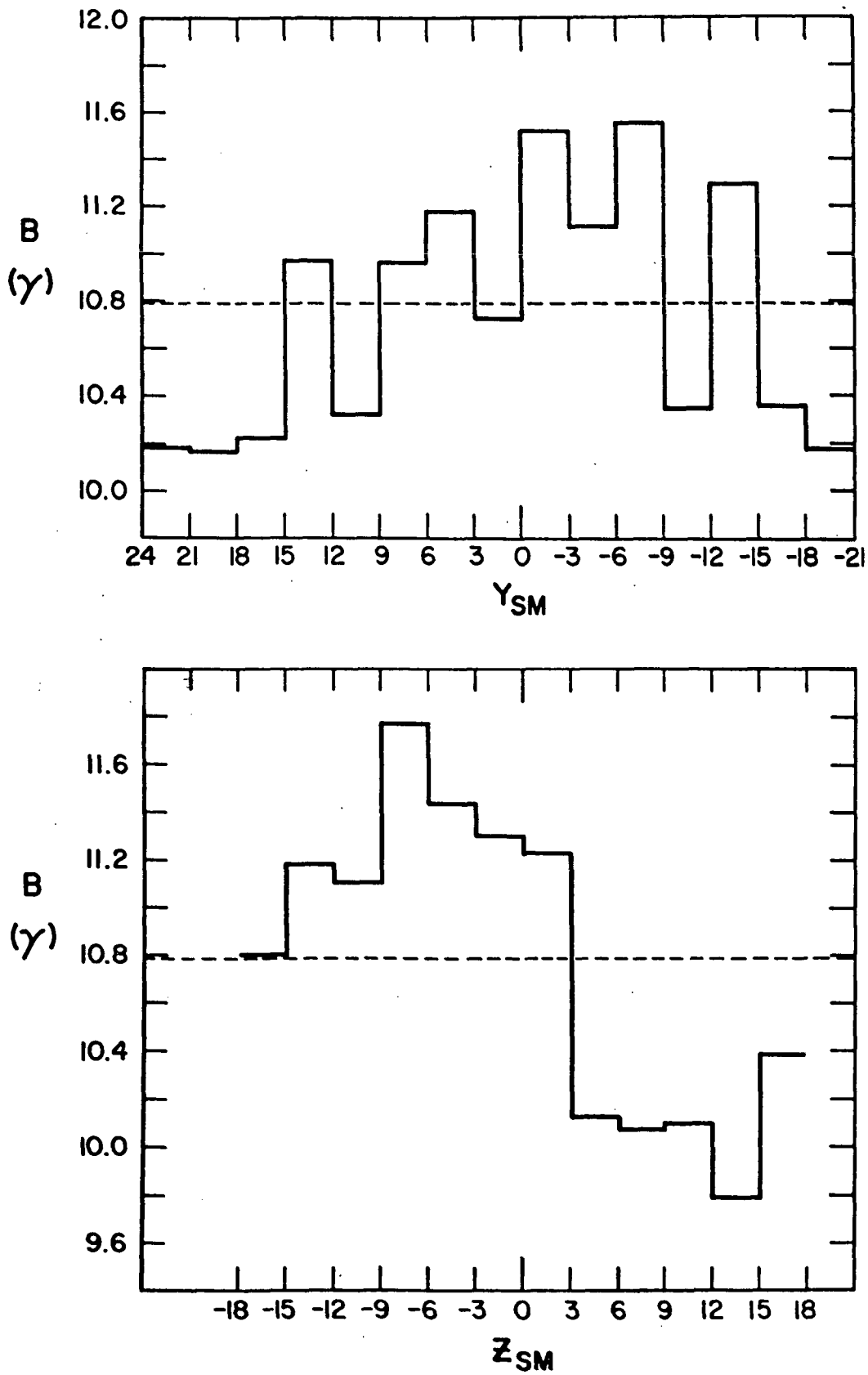

12.0
11.6
11.2
10.8
10.4
10.0
B
(γ)
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
$Y_{SM}$
11.6
11.2
10.8
10.4
10.0
9.6
B
(γ)
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
$Z_{SM}$

Figure 3

# DISTRIBUTIONS OF HIGH LATITUDE LOBE $B_X$, $|B_X|$

Upper No.: $B_X$ Lower No.: $|B_X|$

| $Z_N$ \ $Y_{SM}$ | 24 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 | -3 | -6 | -9 | -12 | -15 | -18 | -21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 10.08<br>*10.08* | 10.93<br>*10.93* | 9.82<br>*9.82* | 9.65<br>*9.65* | | | | | | | | | 11.25<br>*11.25* | 8.09<br>*8.09* | 9.98<br>*9.98* | |
| 15 | 8.92<br>*8.92* | 8.99<br>*8.99* | 11.37<br>*11.37* | 9.36<br>*9.36* | 11.00<br>*11.00* | | | | | | 12.43<br>*12.43* | 6.68<br>*7.07* | 10.26<br>*10.26* | 12.00<br>*12.00* | 9.81<br>*9.81* | |
| 12 | 8.93<br>*8.93* | 8.01<br>*8.01* | 8.98<br>*8.98* | 11.46<br>*11.46* | 8.68<br>*8.68* | 10.55<br>*10.55* | 7.34<br>*7.34* | 8.48<br>*8.48* | 12.19<br>*12.19* | 10.80<br>*10.80* | 10.31<br>*10.31* | 7.79<br>*7.79* | 7.32<br>*7.32* | 11.44<br>*11.44* | 9.11<br>*9.11* | |
| 9 | 9.24<br>*9.24* | 9.27<br>*9.27* | 8.91<br>*8.91* | 10.90<br>*10.90* | 10.23<br>*10.23* | 12.10<br>*12.10* | 10.58<br>*10.58* | 9.34<br>*9.34* | 8.84<br>*8.84* | 9.30<br>*9.30* | 10.39<br>*10.39* | 9.50<br>*9.50* | 9.76<br>*9.76* | 9.65<br>*9.65* | 9.82<br>*9.82* | |
| 6 | 13.46<br>*13.46* | 15.55<br>*15.89* | 11.51<br>*11.51* | 5.09<br>*9.04* | 7.94<br>*10.74* | 7.95<br>*9.46* | 9.94<br>*10.62* | 7.01<br>*7.01* | 8.89<br>*8.89* | 4.51<br>*9.79* | 6.23<br>*9.62* | 10.06<br>*10.06* | 9.88<br>*11.90* | 8.58<br>*8.58* | 12.70<br>*12.70* | |
| 3 | 10.29<br>*12.37* | 8.46<br>*11.21* | 10.69<br>*10.69* | 8.52<br>*8.52* | 7.15<br>*11.06* | -0.15<br>*10.56* | -1.04<br>*11.59* | 4.98<br>*16.13* | 3.22<br>*9.44* | 3.74<br>*8.73* | 7.85<br>*11.63* | 9.57<br>*9.57* | -4.76<br>*11.98* | 9.07<br>*11.25* | -14.30<br>*14.30* | |
| 0 | 9.37<br>*9.37* | -0.68<br>*8.37* | -12.28<br>*16.22* | -4.53<br>*13.16* | -7.15<br>*8.96* | -11.14<br>*12.30* | -10.84<br>*11.66* | -5.76<br>*11.19* | -5.79<br>*13.33* | -5.49<br>*8.87* | -6.14<br>*10.46* | -5.34<br>*11.97* | -7.60<br>*10.49* | -10.56<br>*10.56* | -3.89<br>*3.89* | |
| -3 | -11.06<br>*11.06* | -10.21<br>*10.21* | -7.29<br>*7.29* | -11.08<br>*11.08* | -11.53<br>*11.53* | -10.86<br>*10.86* | -11.14<br>*11.14* | -11.04<br>*11.04* | -11.71<br>*11.71* | -11.95<br>*11.95* | -12.86<br>*12.86* | -10.82<br>*10.82* | -12.17<br>*12.17* | -11.46<br>*11.46* | -12.97<br>*12.97* | |
| -6 | -10.33<br>*10.33* | -7.78<br>*7.78* | -8.94<br>*8.94* | -9.71<br>*9.71* | -9.50<br>*9.50* | -11.25<br>*11.25* | -13.32<br>*13.32* | -13.25<br>*13.25* | -14.27<br>*14.27* | -12.46<br>*12.46* | -12.42<br>*12.42* | -11.19<br>*11.19* | -11.82<br>*11.82* | -10.01<br>*10.01* | -10.01<br>*10.01* | |
| -9 | -8.92<br>*8.92* | -9.97<br>*9.97* | -9.85<br>*9.85* | -11.42<br>*11.42* | -10.98<br>*10.98* | -10.53<br>*10.53* | -10.66<br>*10.66* | -10.34<br>*10.34* | -11.08<br>*11.08* | -15.10<br>*15.10* | -11.96<br>*11.96* | -13.83<br>*13.83* | -11.85<br>*11.85* | -9.20<br>*9.20* | -10.93<br>*10.93* | |
| -12 | -12.80<br>*12.80* | -9.75<br>*9.75* | -9.90<br>*9.90* | -11.82<br>*11.82* | -11.65<br>*11.65* | -10.79<br>*10.79* | | | | | | -11.53<br>*11.53* | -14.17<br>*14.17* | -11.24<br>*11.24* | -13.66<br>*13.66* | |
| -15 | -8.38<br>*8.38* | -10.29<br>*10.29* | -12.06<br>*12.06* | -11.78<br>*11.78* | | | | | | | | | | -10.22<br>*10.22* | | |
| -18 | | | | | | | | | | | | | | | | |

Figure 4

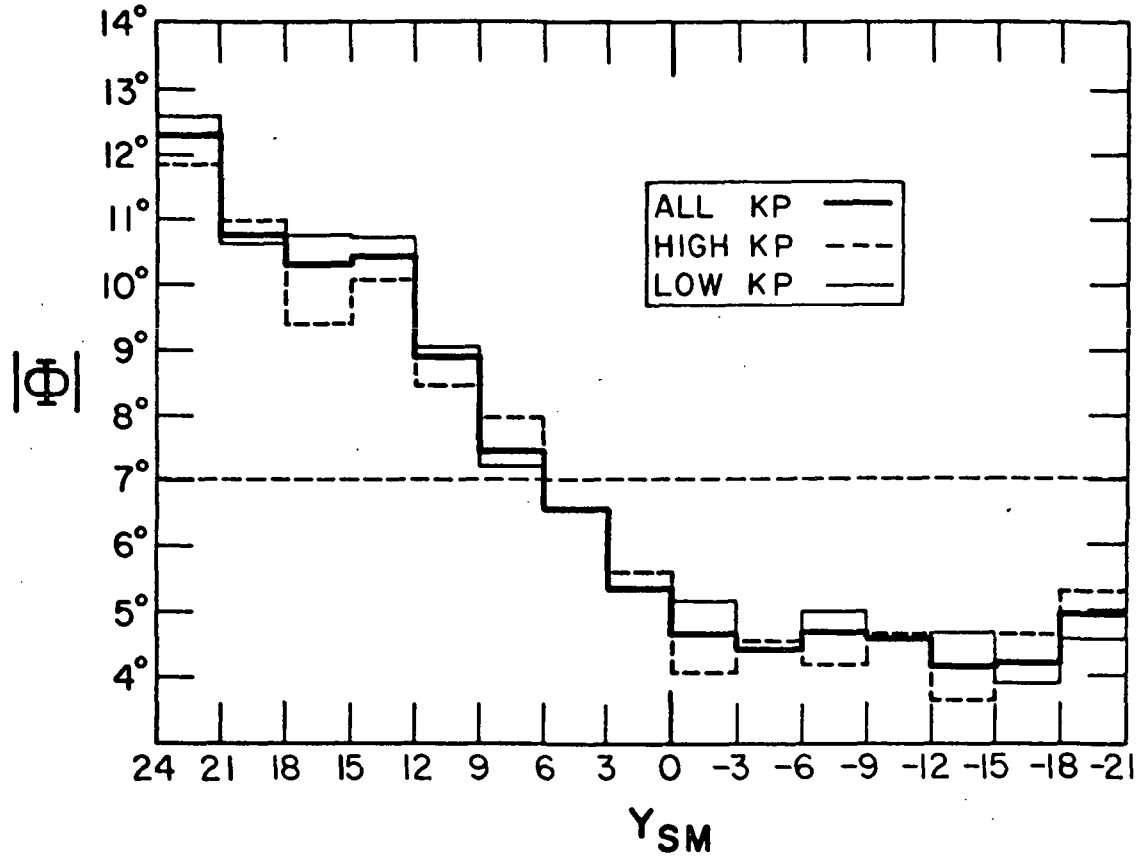
# DISTRIBUTION OF HIGH LATITUDE LOBE $B_Y$

Upper No.: Corrected $B_Y$ ($\gamma$)  Lower No.: Observed $B_Y$ ($\gamma$)

| $Z_N$ | 24 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 0 | -3 | -6 | -9 | -12 | -15 | -18 | -21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | -1.29<br>*-1.89* | -1.87<br>*-2.47* | -1.46<br>*-2.06* | -1.99<br>*-2.59* | | | | | | | | | -0.34<br>*-0.94* | 1.49<br>*0.89* | 0.49<br>*-0.11* | |
| 15 | -1.60<br>*-2.20* | -1.32<br>*-1.92* | -1.91<br>*-2.51* | -1.31<br>*-1.91* | -2.60<br>*-3.20* | | | | | | 1.11<br>*0.51* | 0.05<br>*-0.55* | 0.24<br>*-0.36* | 0.77<br>*0.17* | 1.39<br>*0.79* | |
| 12 | -1.56<br>*-2.16* | -0.87<br>*-1.47* | -1.52<br>*-2.12* | -1.95<br>*-2.55* | -1.21<br>*-1.81* | -1.08<br>*-1.68* | -0.02<br>*-0.62* | 0.12<br>*-0.48* | -0.04<br>*-0.64* | 0.01<br>*-0.59* | 0.08<br>*-0.52* | 0.31<br>*-0.29* | 0.13<br>*-0.47* | -0.71<br>*-1.31* | 0.47<br>*-0.13* | |
| 9 | -1.95<br>*2.55* | -1.28<br>*-1.88* | -0.96<br>*-1.56* | -1.91<br>*-2.51* | -1.35<br>*-1.95* | -1.46<br>*-2.06* | -0.84<br>*-1.44* | -0.48<br>*-1.08* | -0.21<br>*-0.81* | 0.07<br>*-0.53* | -0.25<br>*-0.85* | 0.14<br>*-0.46* | 0.14<br>*-0.46* | 0.74<br>*0.14* | 0.46<br>*-0.14* | |
| 6 | -3.28<br>*-3.88* | -2.90<br>*-3.50* | -2.20<br>*-2.80* | -0.68<br>*-1.28* | -1.11<br>*-1.71* | -0.72<br>*-1.32* | -0.78<br>*-1.38* | 0.42<br>*-0.18* | 0.73<br>*0.13* | 0.26<br>*-0.34* | 0.22<br>*-0.38* | -0.02<br>*-0.62* | 0.33<br>*-0.27* | 0.49<br>*-0.11* | 0.72<br>*0.12* | |
| 3 | -1.35<br>*-1.95* | -1.05<br>*-1.65* | -1.99<br>*-2.59* | -1.08<br>*-1.68* | 0.35<br>*-0.25* | 0.09<br>*-0.51* | 0.91<br>*0.31* | -0.53<br>*-1.13* | 0.33<br>*-0.27* | 0.19<br>*-0.41* | -0.04<br>*-0.64* | -0.15<br>*-0.75* | 0.26<br>*-0.34* | 0.78<br>*0.18* | 0.21<br>*-0.39* | |
| 0 | -2.10<br>*-1.50* | -0.52<br>*0.08* | 2.45<br>*3.05* | 0.17<br>*0.57* | 0.52<br>*0.92* | 0.55<br>*0.95* | 0.31<br>*0.71* | -0.08<br>*0.52* | -0.09<br>*0.51* | -0.76<br>*-0.16* | -0.85<br>*-0.25* | -1.57<br>*-0.97* | -1.27<br>*-0.67* | -0.53<br>*0.07* | -0.63<br>*-0.03* | |
| -3 | 1.65<br>*2.25* | 1.18<br>*1.78* | 0.50<br>*1.10* | 1.43<br>*2.03* | 0.73<br>*1.33* | 0.37<br>*0.97* | -0.53<br>*0.07* | -0.03<br>*0.37* | -0.17<br>*0.43* | -0.23<br>*0.37* | -0.23<br>*0.37* | -1.18<br>*-0.58* | -1.37<br>*-0.97* | -0.92<br>*-0.32* | -2.60<br>*-2.00* | |
| -6 | 1.02<br>*1.62* | 0.49<br>*1.09* | 1.20<br>*1.80* | 0.71<br>*1.31* | 0.26<br>*0.86* | 0.28<br>*0.88* | 0.18<br>*0.78* | 0.36<br>*0.96* | -0.25<br>*0.35* | -0.61<br>*-0.01* | -0.74<br>*-0.14* | -1.12<br>*-0.52* | -0.89<br>*-0.29* | -1.27<br>*-0.67* | -1.25<br>*-0.65* | |
| -9 | 0.81<br>*1.41* | 1.40<br>*2.00* | -0.06<br>*0.54* | 0.50<br>*1.10* | 0.35<br>*0.95* | 0.38<br>*0.98* | -0.10<br>*0.50* | -0.42<br>*0.18* | -1.66<br>*-1.06* | -0.38<br>*0.22* | -0.92<br>*-0.32* | -1.40<br>*-0.80* | -0.87<br>*-0.27* | -1.14<br>*-0.54* | -1.00<br>*-0.40* | |
| -12 | 0.47<br>*1.07* | 0.79<br>*1.39* | 0.73<br>*1.33* | 1.17<br>*1.77* | 0.68<br>*1.28* | 0.73<br>*1.33* | | | | | | -1.39<br>*-0.79* | -1.53<br>*-0.93* | -1.33<br>*-0.73* | -1.77<br>*-1.17* | |
| -15 | 0.30<br>*0.90* | 1.05<br>*1.65* | 1.20<br>*1.80* | 1.48<br>*2.08* | | | | | | | | | | -0.87<br>*-0.27* | | |
| -18 | | | | | | | | | | | | | | | | |

$Y_{SM}$

Figure 5


14°
13°
12°
11°
10°
9°
8°
7°
6°
5°
4°
|Φ|
ALL KP
HIGH KP
LOW KP
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
YSM

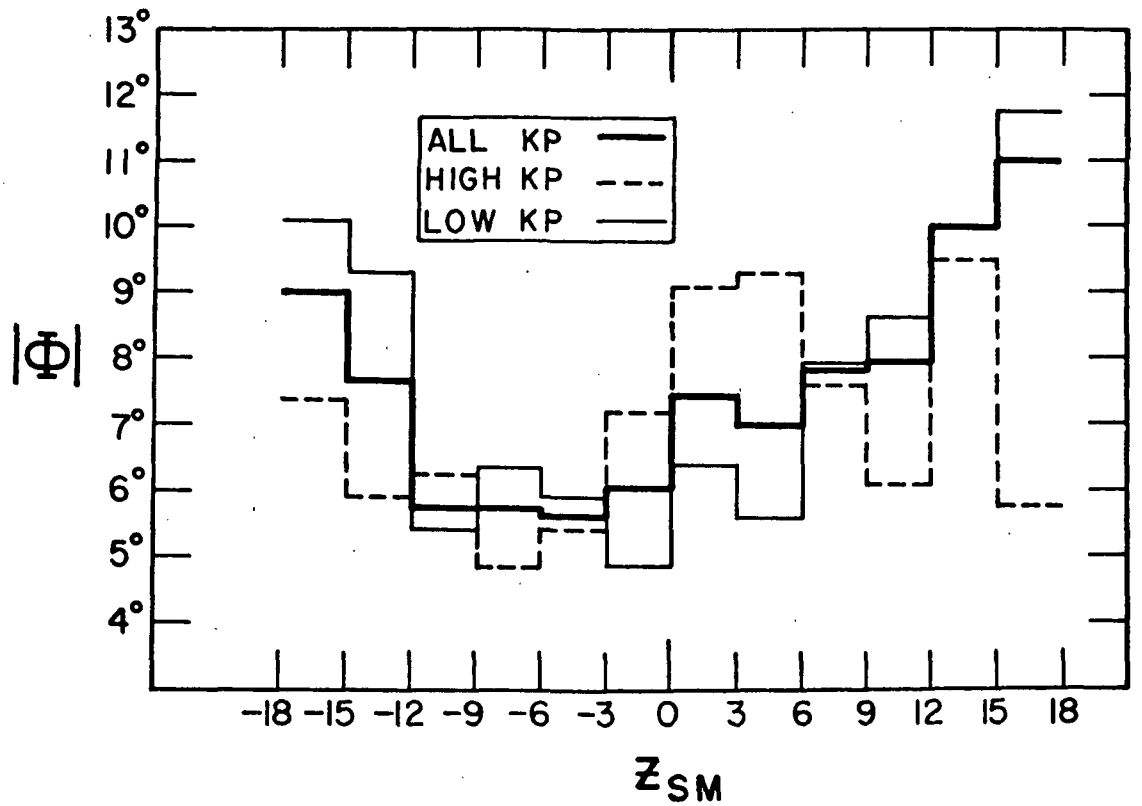
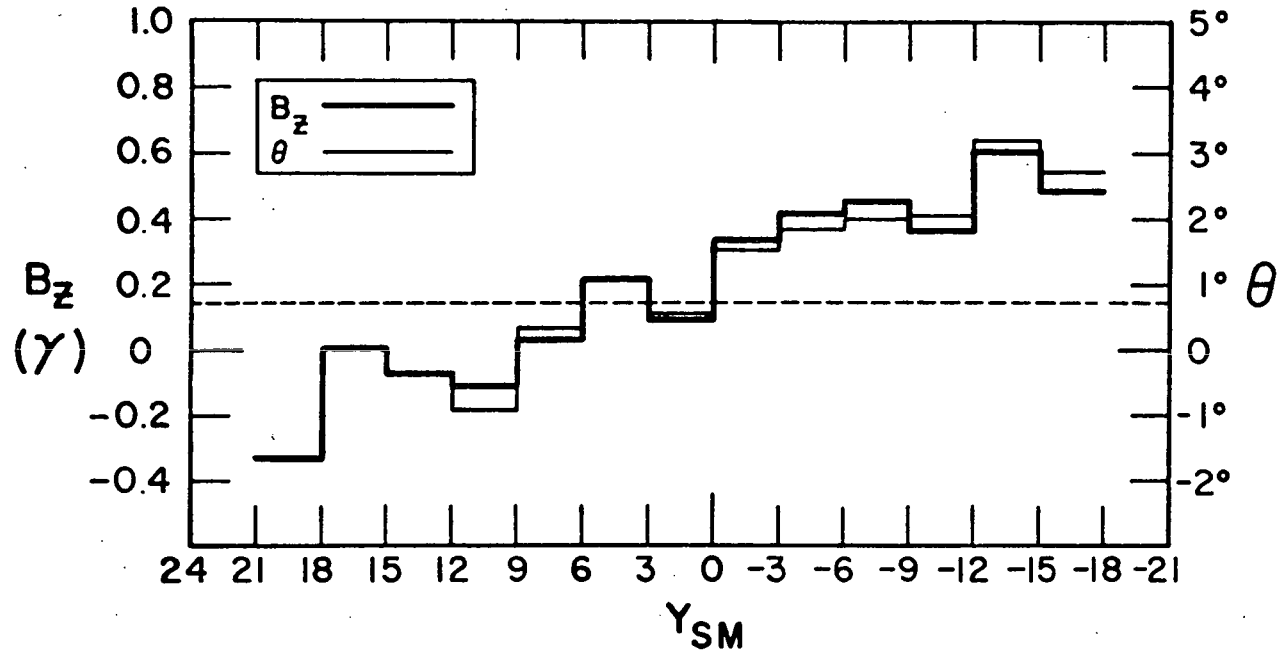

13°
12°
11°
10°
9°
8°
7°
6°
5°
4°
|Φ|
ALL KP
HIGH KP
LOW KP
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
ZSM

Figure 6


1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
$B_z$
($\gamma$)
$B_z$
$\theta$
5°
4°
3°
2°
1°
0
-1°
-2°
$\theta$
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
$Y_{SM}$

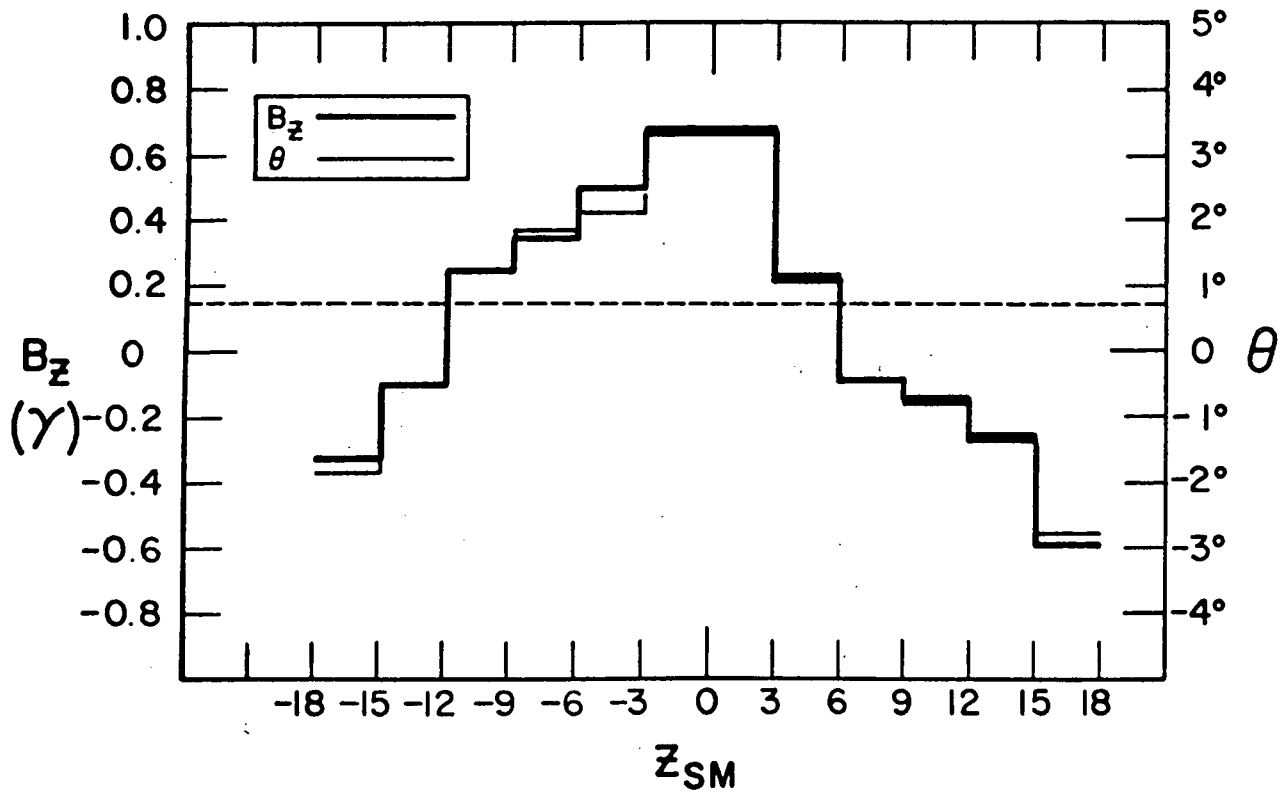

1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
$B_z$
($\gamma$)
$B_z$
$\theta$
5°
4°
3°
2°
1°
0
-1°
-2°
-3°
-4°
$\theta$
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
$Z_{SM}$

Figure 7

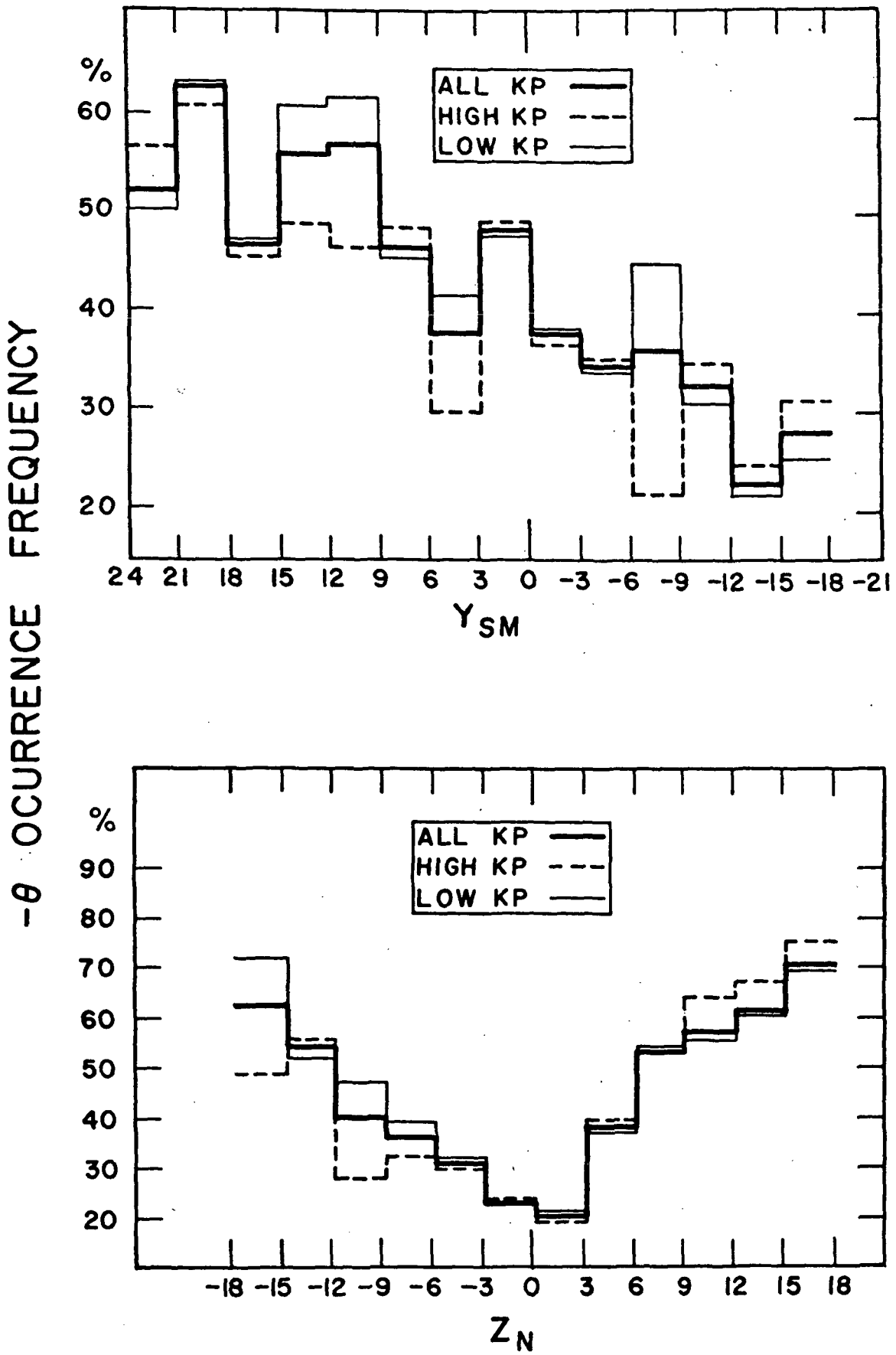

-θ OCURRENCE FREQUENCY
%
60
50
40
30
20
ALL KP
HIGH KP
LOW KP
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
YSM
%
90
80
70
60
50
40
30
20
ALL KP
HIGH KP
LOW KP
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
ZN

Figure 8

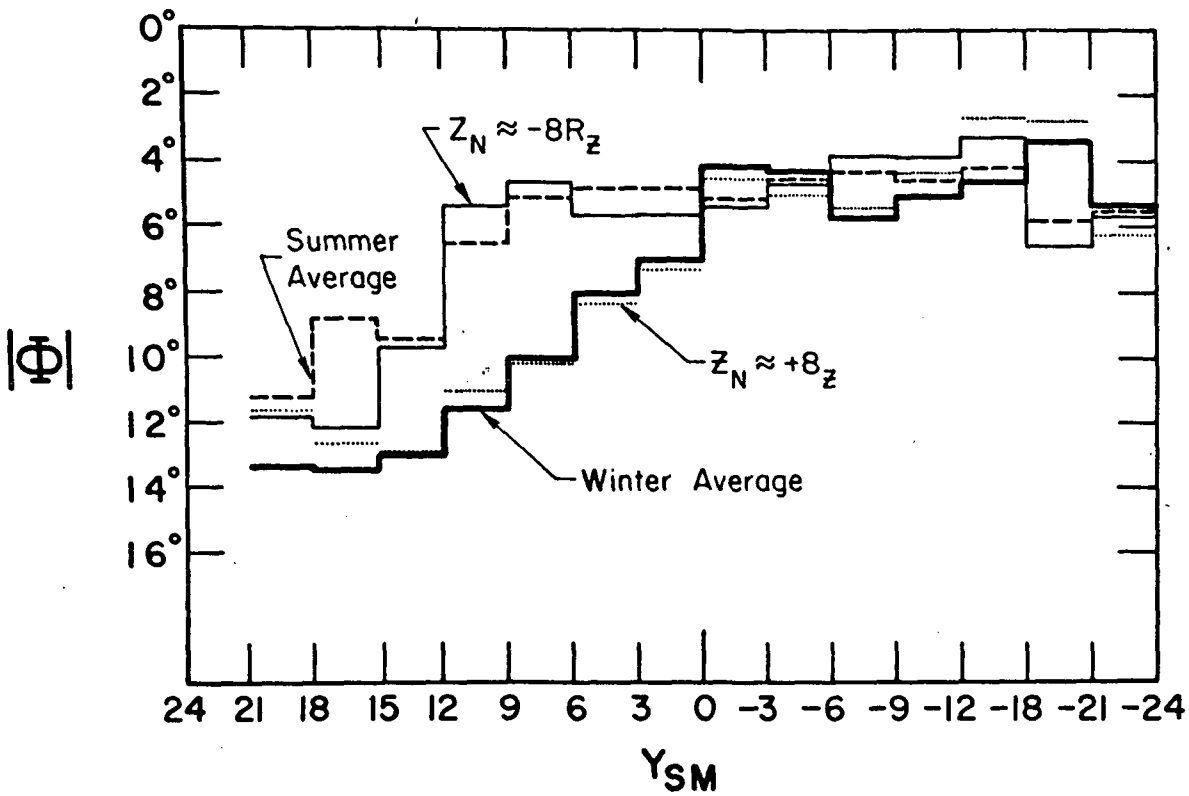

0°
2°
4°
6°
8°
10°
12°
14°
16°
|Φ|
$Z_N \approx -8R_z$
Summer Average
$Z_N \approx +8_z$
Winter Average
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -18 -21 -24
$Y_{SM}$

Figure 9

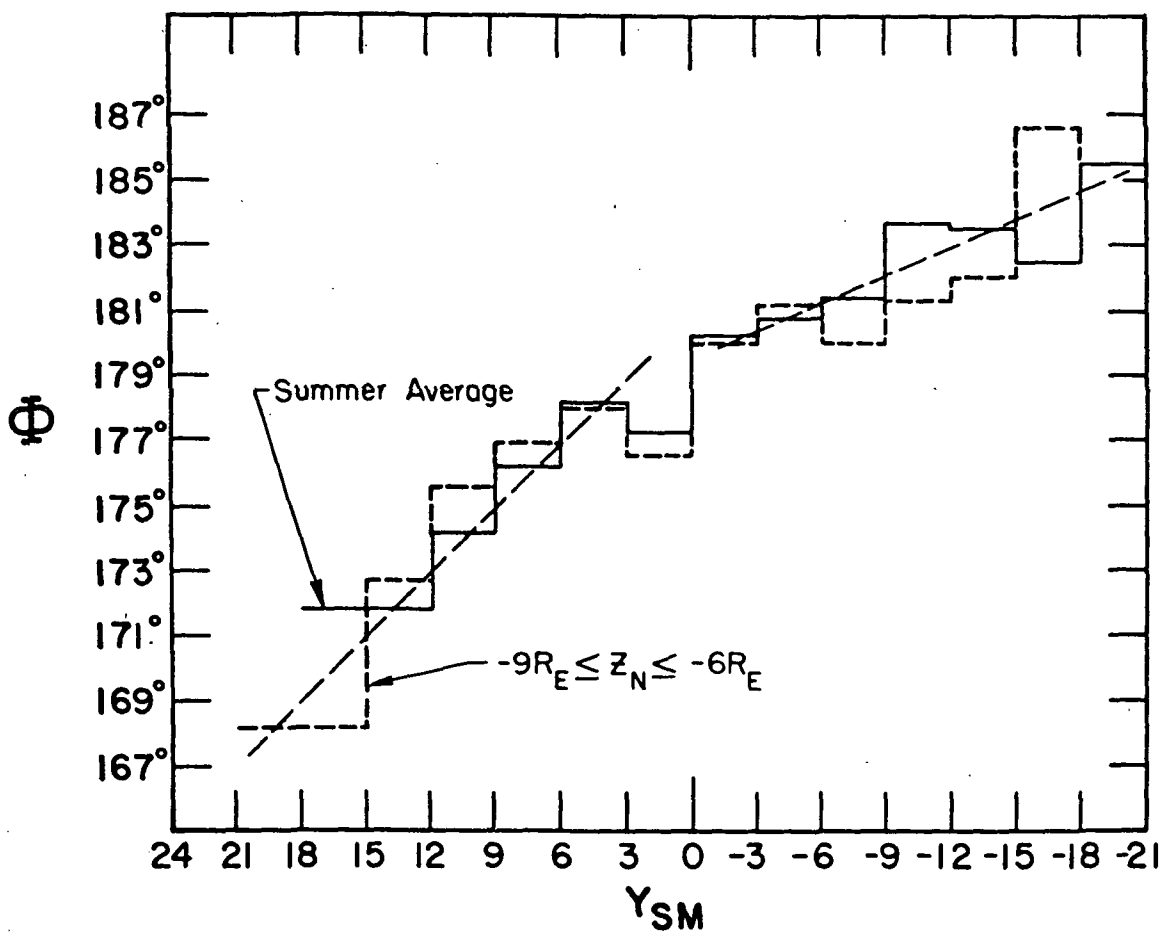

Φ
187°
185°
183°
181°
179°
177°
175°
173°
171°
169°
167°
Summer Average
$-9R_E \leq Z_N \leq -6R_E$
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
$Y_{SM}$

Figure 10

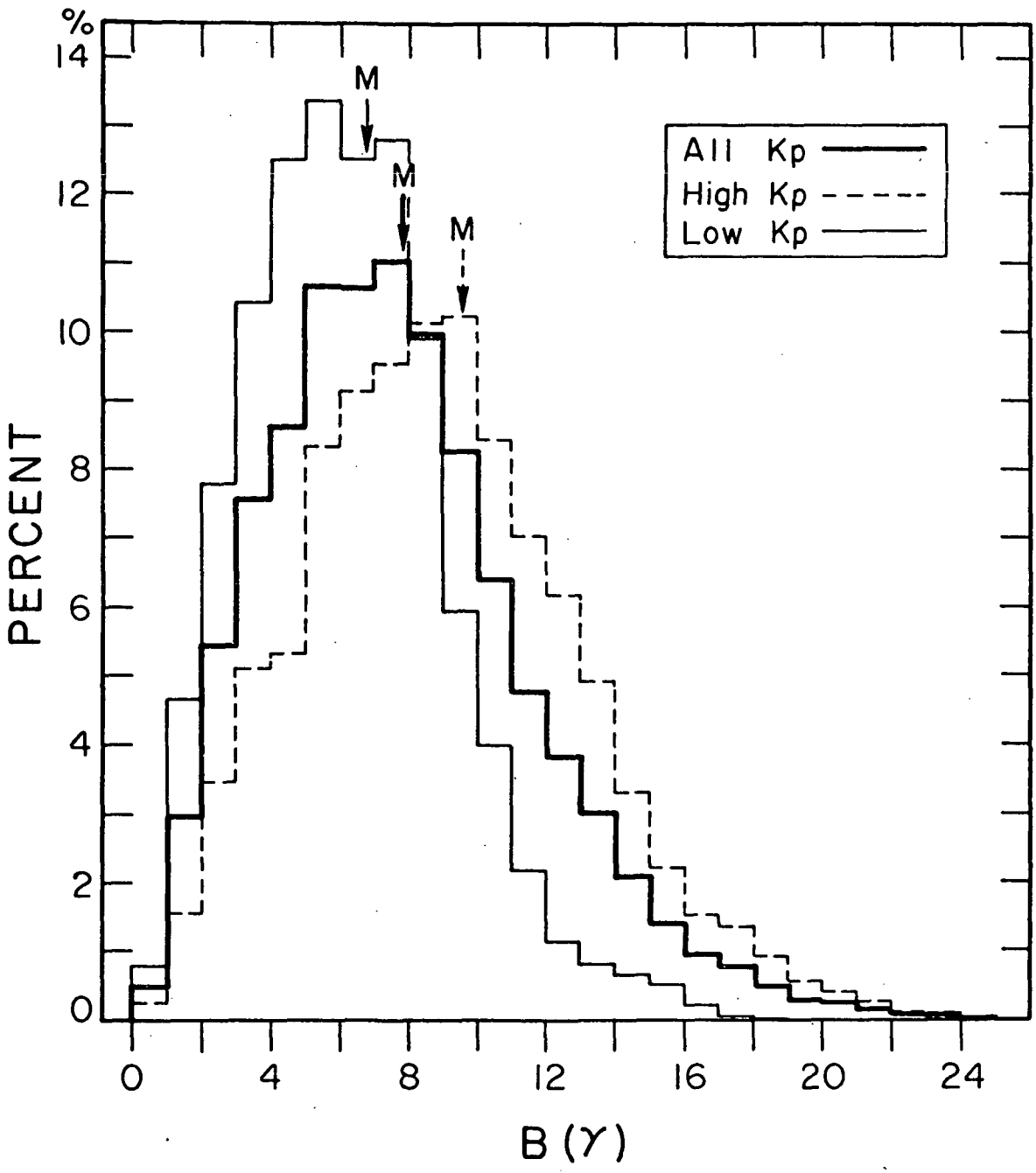

%
14
12
10
8
6
4
2
0
PERCENT
M
M
M
All Kp
High Kp
Low Kp
0
4
8
12
16
20
24
B (γ)

Figure 11

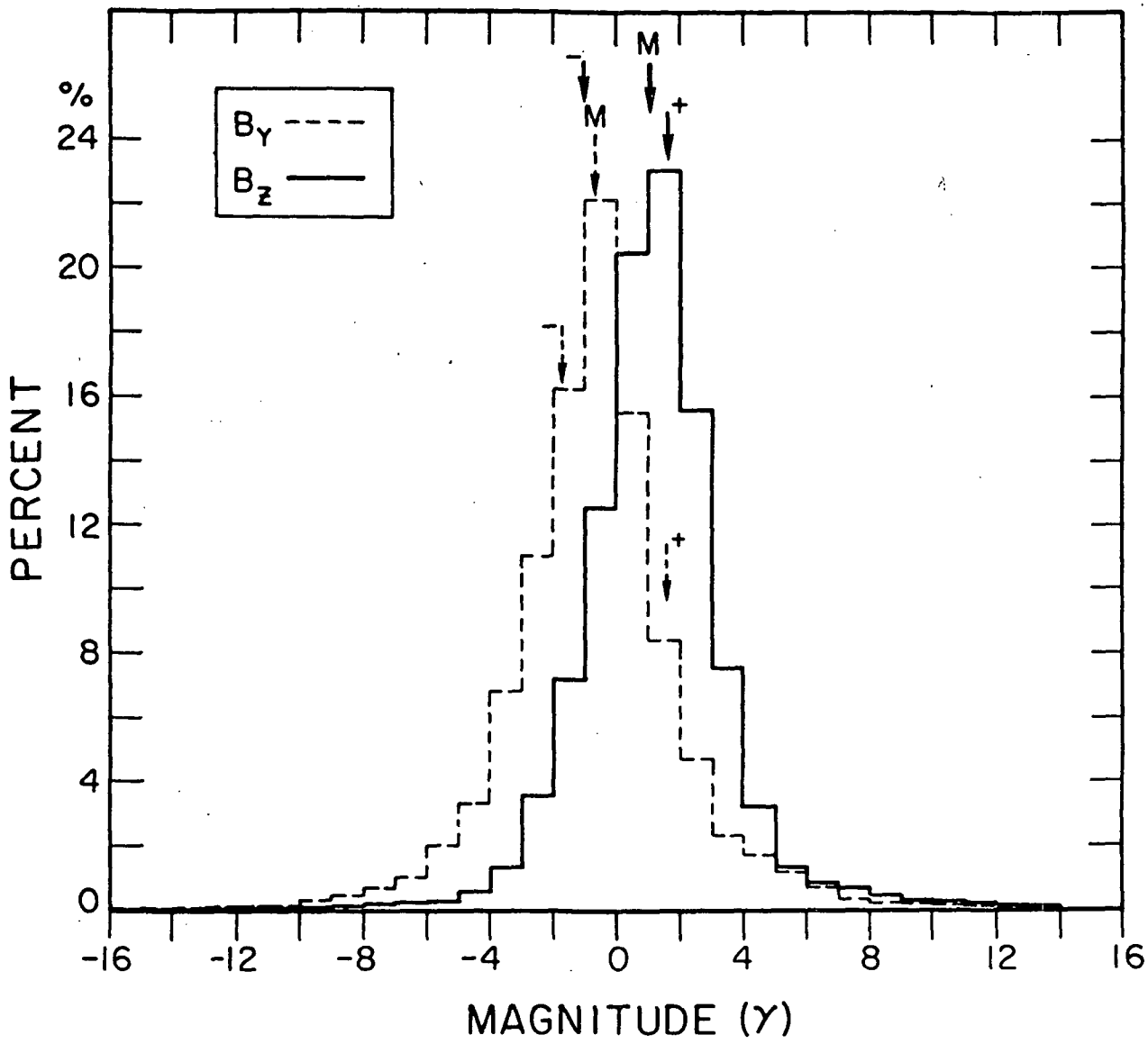



Figure 11
(cont.)

# DISTRIBUTIONS OF PLASMA SHEET $B_x$, $|B_x|$

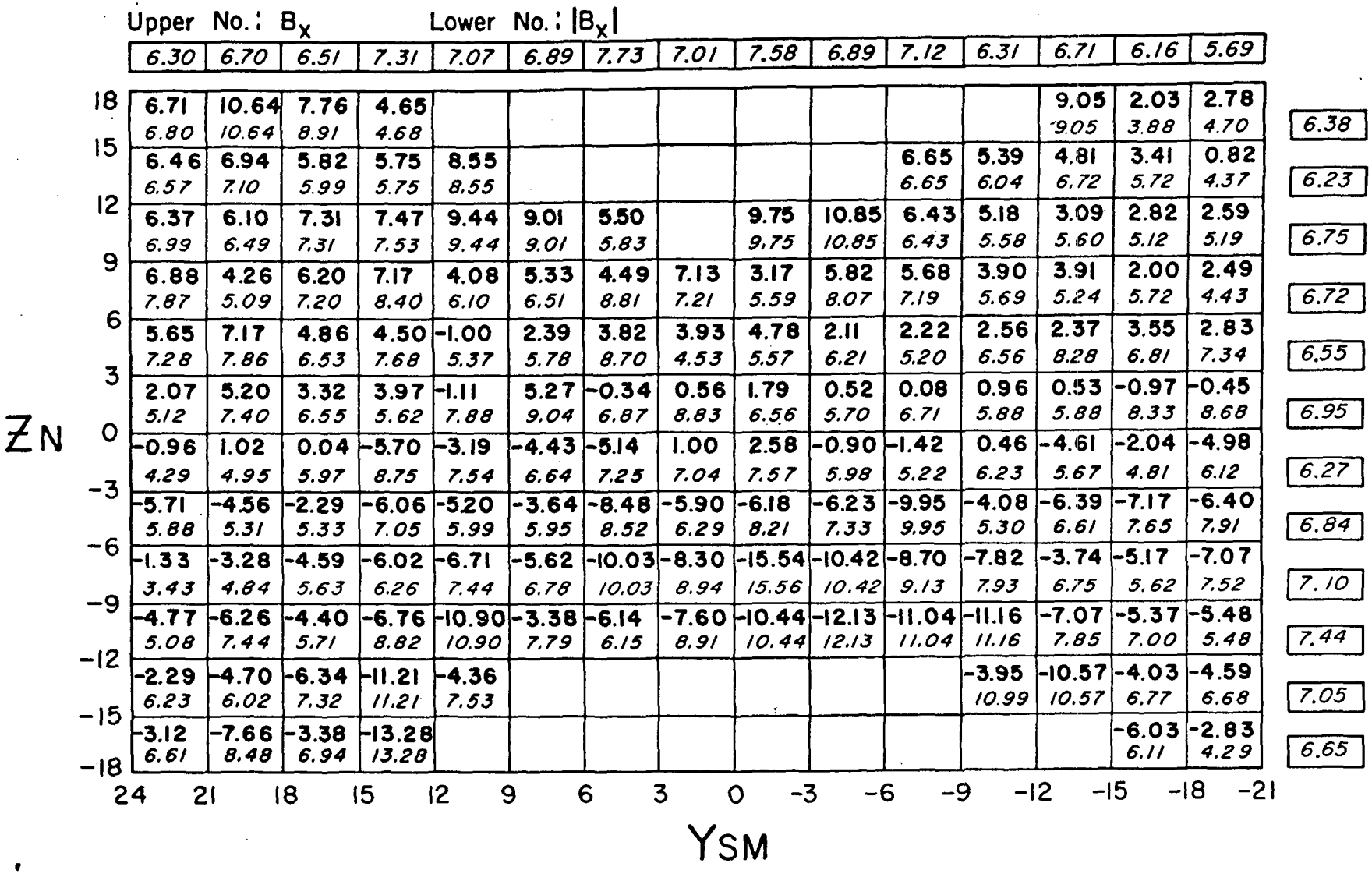
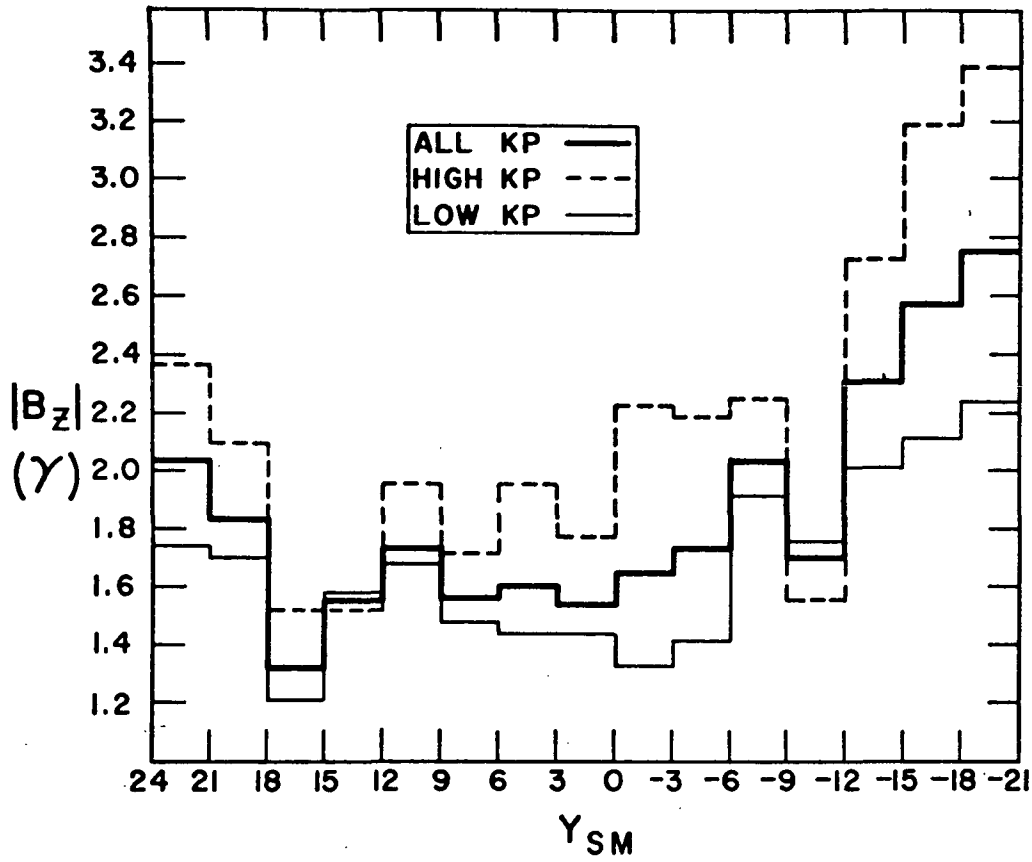



Figure 12


ALL KP
HIGH KP
LOW KP
|Bz| (γ)
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
YSM

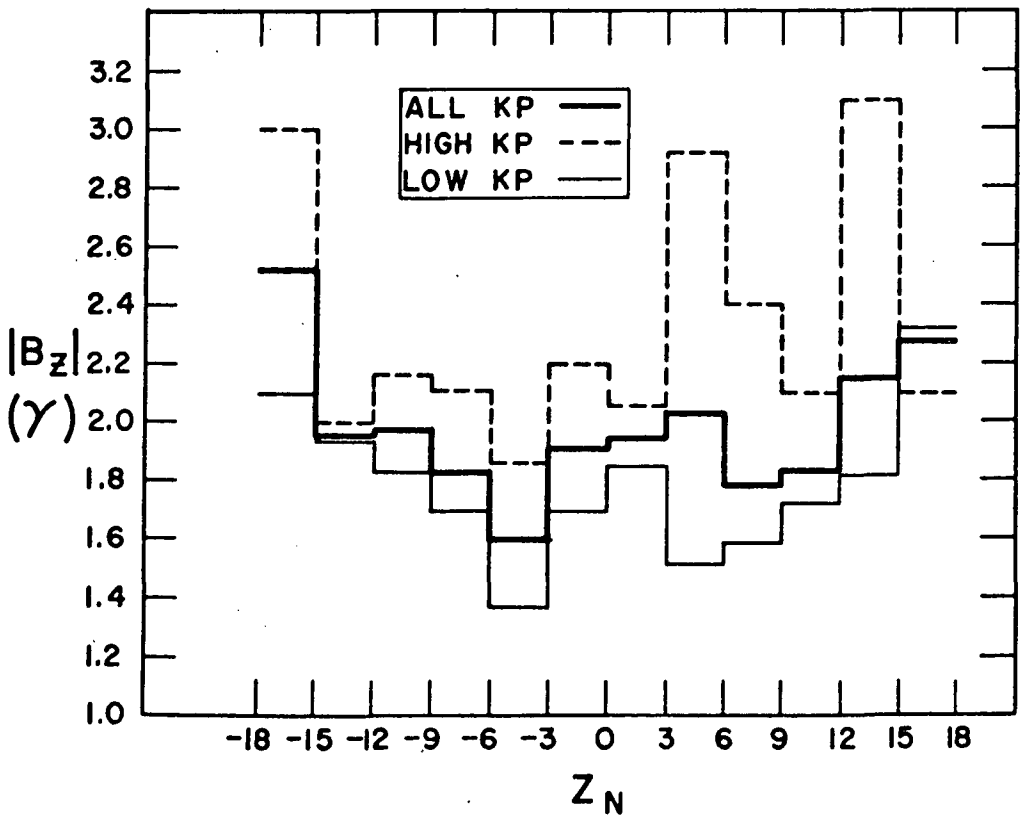

ALL KP
HIGH KP
LOW KP
|Bz| (γ)
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
ZN

Figure 13

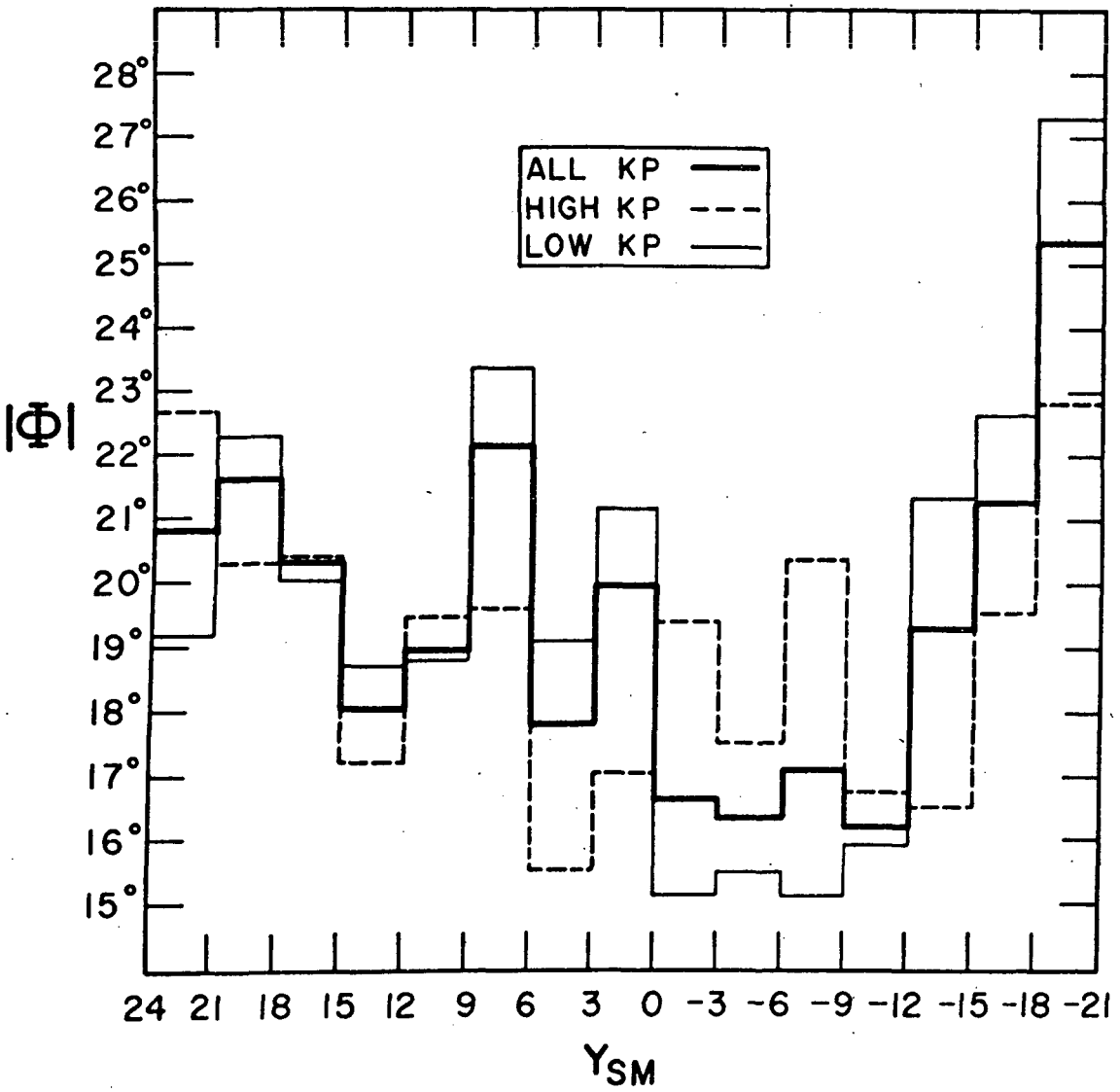

ALL KP
HIGH KP
LOW KP
|Φ|
28°
27°
26°
25°
24°
23°
22°
21°
20°
19°
18°
17°
16°
15°
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
YSM

Figure 14

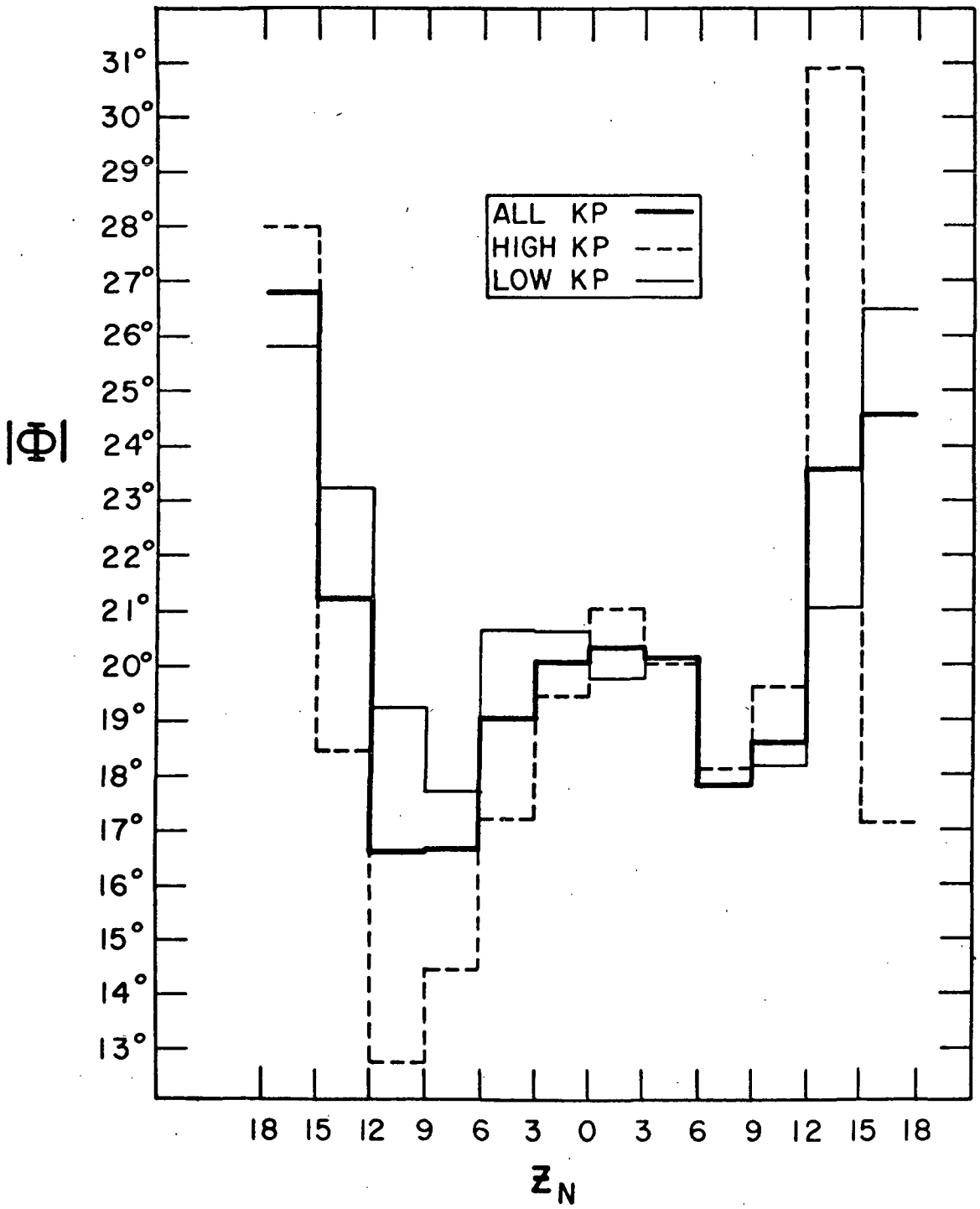
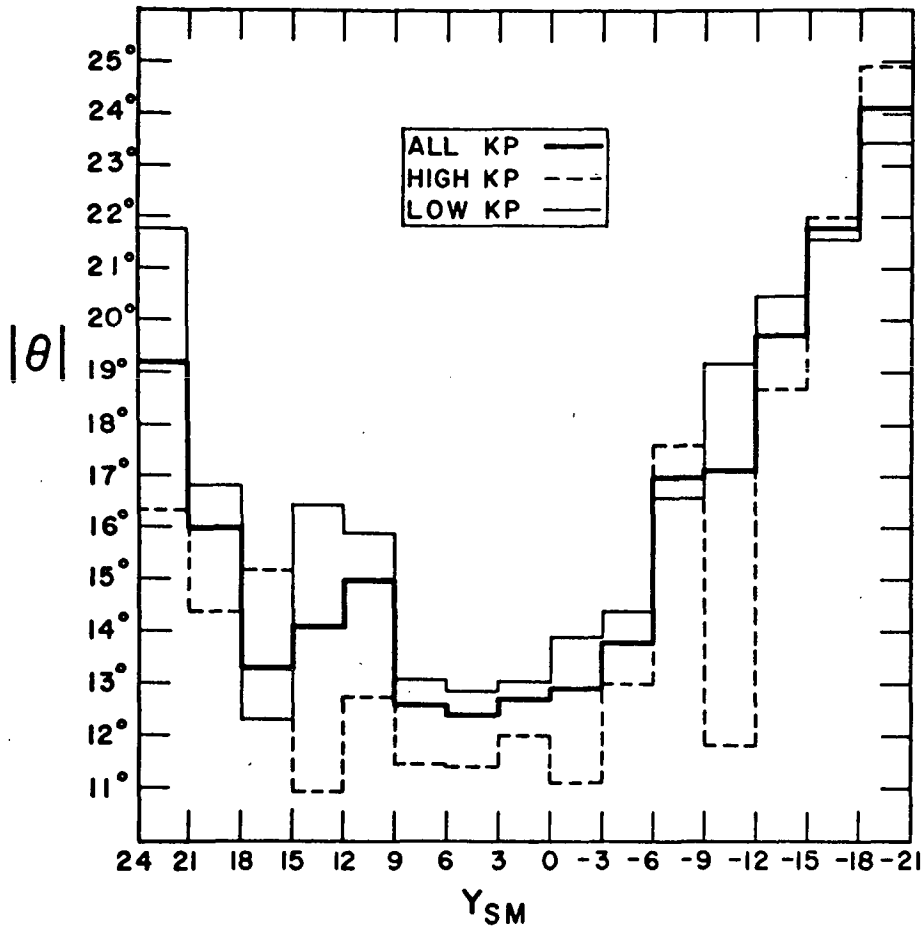



Figure 14
(cont.)


ALL KP
HIGH KP
LOW KP
|θ|
25°
24°
23°
22°
21°
20°
19°
18°
17°
16°
15°
14°
13°
12°
11°
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
Y_SM

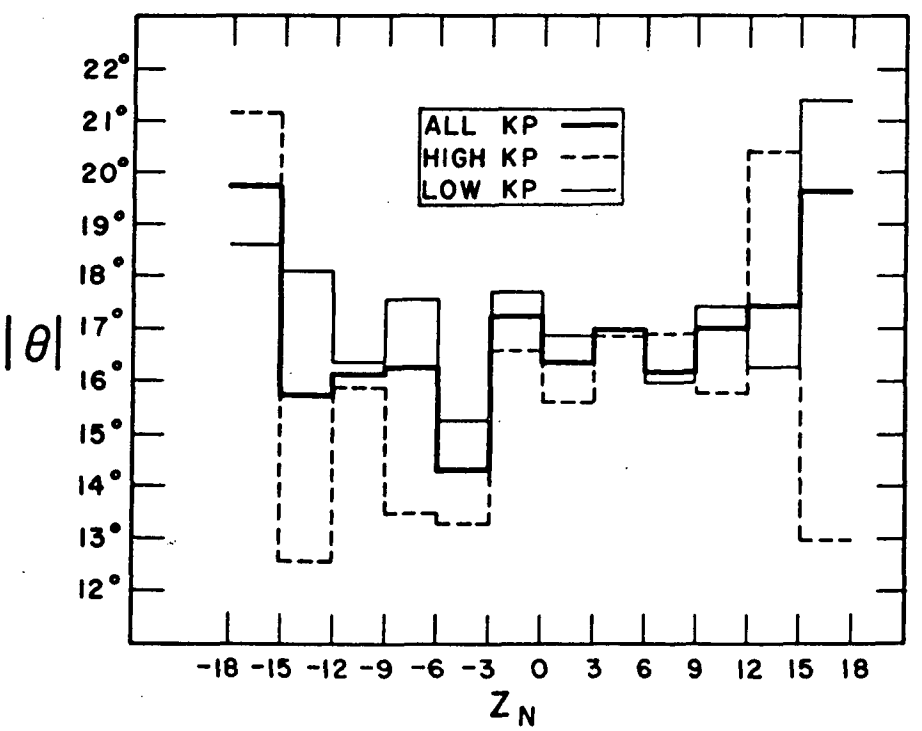

ALL KP
HIGH KP
LOW KP
|θ|
22°
21°
20°
19°
18°
17°
16°
15°
14°
13°
12°
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
Z_N

Figure 15

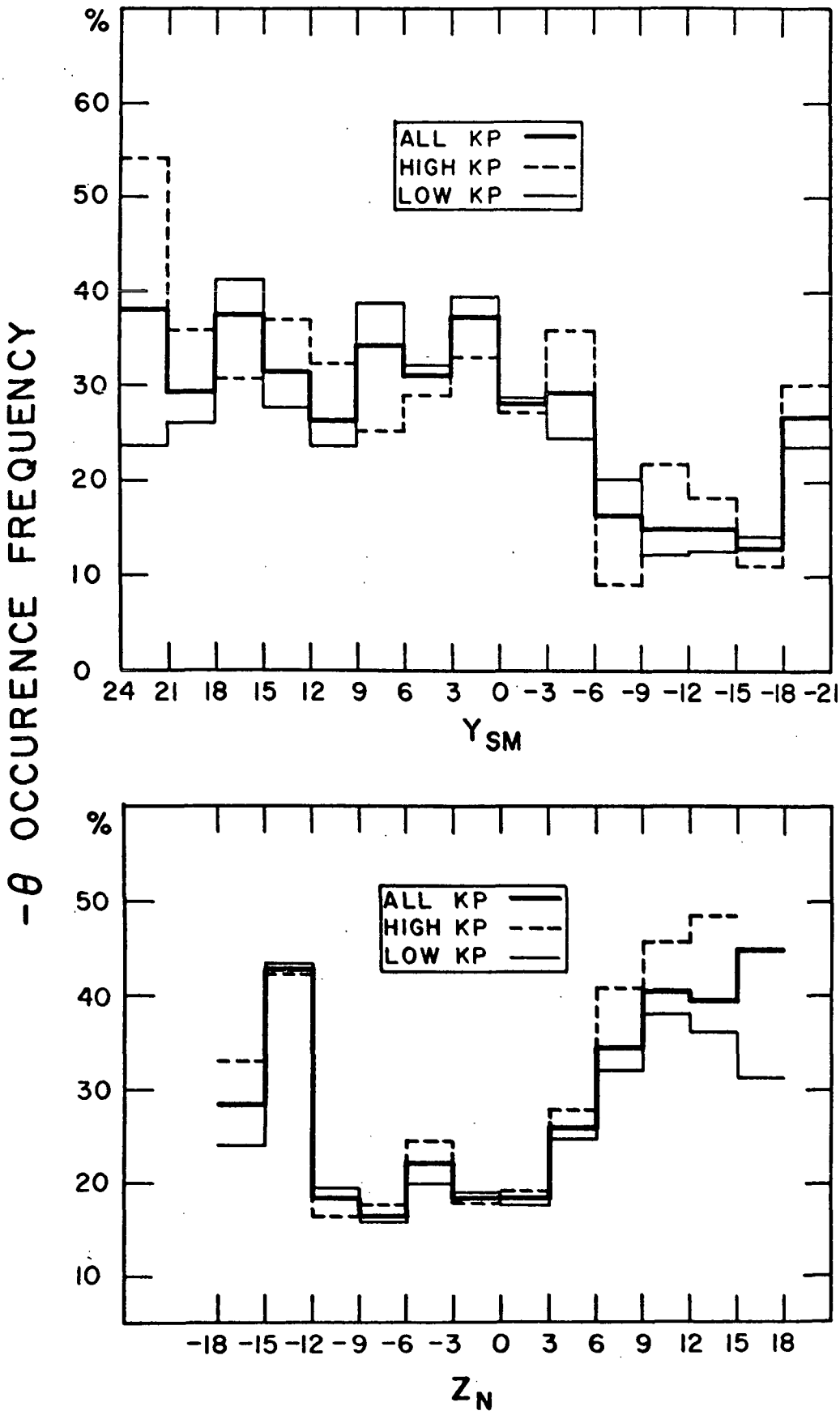

-θ OCCURENCE FREQUENCY
%
60
50
40
30
20
10
0
ALL KP
HIGH KP
LOW KP
24 21 18 15 12 9 6 3 0 -3 -6 -9 -12 -15 -18 -21
Y SM
%
50
40
30
20
10
ALL KP
HIGH KP
LOW KP
-18 -15 -12 -9 -6 -3 0 3 6 9 12 15 18
Z N

Figure 16

A74-14371

# On the Distribution of $He^+$ in the Plasmasphere From Observations of Resonantly Scattered He II 304-Å Radiation

FRANCESCO PARESCE, C. STUART BOWYER, AND SHAILENDRA KUMAR[1]

*Space Sciences Laboratory, University of California, Berkeley, California 94720*

The intensity of the He II 304-Å radiation in the night sky was measured from a sounding rocket launched from White Sands Missile Range on June 9, 1972. The data obtained on this flight are compared with predictions of theoretical models of the $He^+$ altitude distribution within the plasmasphere. The observations are not consistent with an isothermal equilibrium model of the helium ion in the plasmasphere, confirming earlier results. The data are best fit by a diffusive equilibrium model of exospheric $He^+$ having a plasma temperature gradient of 0.5°K $km^{-1}$ and a number density at 1600 km of $2 \times 10^5$ $He^+$ ions $cm^{-3}$. In this model the ion densities are calculated without allowance for thermal diffusion. A similar model that allows for thermal diffusion and that has a $He^+$ number density at 1600 km of $1.5 \times 10^3$ $cm^{-3}$ and a temperature gradient of 1.2°K $km^{-1}$ yields a less satisfactory fit to the observations but is more consistent with in situ satellite observations. An empirical power law model in which the ion density decreases as $R^{-3}$ from a value of $6 \times 10^3$ $He^+$ ions $cm^{-3}$ at 1600 km is marginally consistent with the observations.

The existence of a substantial flux (≃1–10 R) of He II 304-Å radiation in the night sky at rocket altitudes has been well established by photometric techniques [*Young et al.*, 1971; *Ogawa and Tohmatsu*, 1971; *Paresce et al.*, 1973]. The observed intensities and spatial variations can be interpreted as being primarily due to resonant scattering of solar He II 304-Å radiation by singly ionized helium in the illuminated plasmasphere above the rocket. In a small region around the antisolar direction, extraplasmaspheric 304-Å radiation may be playing some as yet unclear role. Satisfactory fits to the available data in view directions excluding the antisolar region have been obtained with a simple plasmaspheric model in which the $He^+$ density is assumed constant throughout the plasmasphere, an average value of a few hundred ions per cubic centimeter being used for the $He^+$ density [*Meier and Weller*, 1972; *Paresce et al.*, 1973]. These results are consistent with a number of observational results and theoretical predictions but are inconsistent with others. There is some uncertainty about whether the discrepancies encountered are real or are due to experimental difficulties.

To obtain additional data on this question, two extreme ultraviolet photometers were flown to an altitude of 206 km on a Black Brant VC launched from White Sands Missile Range at 0525 UT on June 9, 1972. At the time of this flight the sun was in the north, 35° below the horizon. The purpose of this paper is to report observations made with one of these instruments on the intensity of the He II 304-Å night glow and to interpret the data obtained in terms of plasmaspheric $He^+$ models.

## Instrumentation and Flight

The measurements were made using a 1-cm-diameter funnel-mouthed channel electron multiplier detector with a thin metallic film filter consisting of 1000 Å of aluminum overcoated with 300 Å of carbon. The combination of photocathode sensitivity and filter transmission restricted the band pass of the instrument to the range approximately 170–550 Å. Instrumentation of this type has been described in detail by *Bowyer et al.* [1970]. The detector employed in this study is identical to that described by *Paresce et al.* [1973], and we refer the reader to that paper for details on the calibration and particle rejection techniques employed. The specific characteristics of the photometer used in this study are as follows: filter, 1000 Å of Al + 300 Å of C; bandpass (to 10% transmission), 170–550 Å; field of view, 7° full width at half maximum; effective area, 0.8 $cm^2$; effective solid angle, 0.012 sr; collimator transmission, 0.54; quantum efficiency (counts/photon) at 304 Å, $7.6 \times 10^{-3}$; 584 Å, $8.2 \times 10^{-4}$, 1216 Å, $4.5 \times 10^{-7}$. Another photometer of the same type but using a thin film filter consisting of 1600 Å of tin was flown on the same rocket to obtain information on He I 584-Å radiation that may contaminate the measurements of the He II line. Data on this line will be published in a subsequent paper.

The detector was mounted looking out the side of the rocket with a view perpendicular to the spin axis. During the flight the rocket motion was controlled by an attitude control system, and the aspect was recorded by a 35-mm camera that photographed star fields every 1.78 s. The aspect solution derived in this manner is accurate to ±6 arc min. In Figure 1*a* we show the rocket altitude in kilometers as a function of time from launch. We also show the height in earth radii from the center of the earth of the intersection of the line of sight of the photometer with the shadow cylinder of the earth at 304 Å. This cylinder is assumed to graze the earth at an altitude of 300 km above the earth's surface, an altitude at which a grazing incidence ray at 304 Å suffers an $e^{-1}$ attenuation. The zenith angle and azimuth of the center of the field of view of the detector throughout the flight are shown in Figure 1*b*.

[1] Now at Kitt Peak National Observatory, Tucson, Arizona 85717.

Copyright © 1974 by the American Geophysical Union.

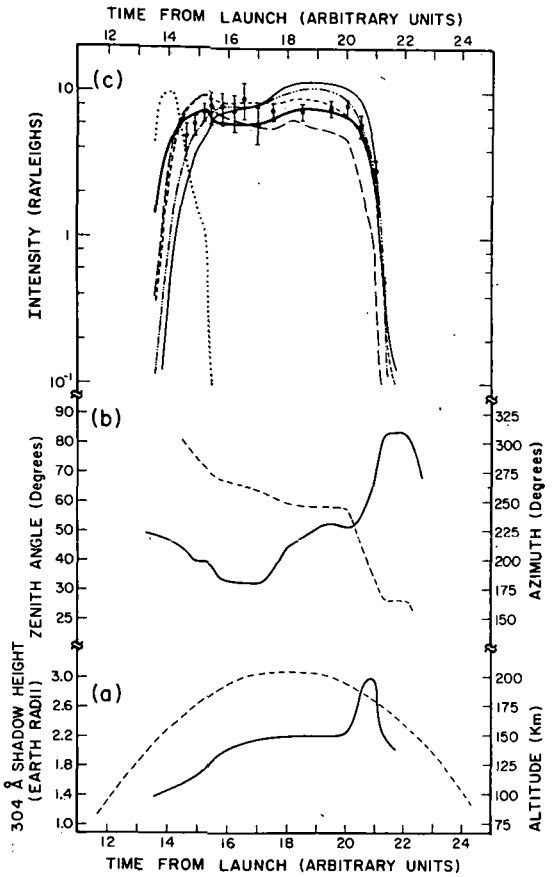



Fig. 1. (*a*) Rocket altitude from the earth's surface (dashed line) and 304-Å shadow height at the center of the photometer's field of view from the center of the earth as a function of time from launch (solid line). (*b*) Azimuth of the center of the detector's field of view measured clockwise from the north (dashed line) and zenith angle of the center of the detector's field of view as a function of time from launch (solid line). (*c*) Observed intensities as a function of time from launch (filled circles), predicted intensities using isothermal diffusive equilibrium model (dotted line), constant temperature gradient diffusive equilibrium model (thick solid line), constant density model (faint solid line), $R^{-4}$ power model (long-dashed line), $R^{-3}$ power law model (short-dashed line), and diffusive equilibrium with thermal diffusion model (dash-dot line).

## Results and Interpretation

An altitude and aspect dependent signal was recorded by the photometer above approximately 140 km. Below this altitude on both legs of the flight a constant background of a few counts per second was recorded, which we believe to be due to slight Lyman $\alpha$ leakage through pinholes in the filter. This count rate was subtracted from the data obtained above 140 km. The data were also corrected for a small residual contribution due to He I 584-Å radiation observed in the same flight by the tin filter detector. *Carlson* [1972] has recently pointed out that ionospheric $O^{++}$ can also scatter the solar He II 304-Å line and consequently possibly contribute to the observations. This source would indeed be important if the measurements were made in daytime or if the intersection of the line of sight with the earth's shadow cylinder occurred at altitudes below 1000 km, where the $O^{++}$ and $He^{+}$ densities may be comparable. At the altitudes of our observations the $O^{++}$ density is about 2 orders of magnitude less than the $He^{+}$ density [*Hoffmann*, 1970], however, and the contribution of this source to the measured signals is therefore negligible. If it is assumed that there are no other emissions in the detector's band pass, the resultant He II 304-Å emission rates may be computed, and they are shown in Figure 1*c* as filled circles. The errors indicated are those associated with the statistical fluctuations in the measurements only. The peak intensity observed was $9 \pm 2$ R at an altitude of 200 km and at a zenith angle of $\simeq 32°$. These results are consistent with previous measurements of this radiation [*Young et al.*, 1971; *Ogawa and Tohmatsu*, 1971; *Paresce et al.*, 1973].

Since the scattering medium is optically thin [*Paresce et al.*, 1973], the intensity $I$ of the 304-Å line in rayleighs is given by the relation

$$I = 10^{-6} \times A \int_s \int_\nu \pi F_\nu a_\nu p(\theta) n(\mathrm{He}^+)\, ds\, d\nu \qquad (1)$$

where the integrations are to be performed along the line of sight $s$ and over frequency $\nu$. In this expression, $\pi F_\nu$ is the solar 304-Å line flux, $a_\nu$ is the scattering cross section, $p(\theta)$ is the phase function for scattering through an angle $\theta$, and $n(\mathrm{He}^+)$ is the number density of $He^+$ at any point on the integration path. The attenuation coefficient due to photoabsorption of 304-Å radiation by $O_2$, O, and $N_2$ above the rocket is indicated by $A$. The form of $p(\theta)$ for this case is given by the relation [*Brandt and Chamberlain*, 1959]

$$p(\theta) = 1 + \tfrac{1}{4}(\tfrac{2}{3} - \sin^2\theta) \qquad (2)$$

The integration path may be separated into at least three distinct sections: the plasmasphere, the outer magnetosphere, and the interplanetary medium. The contribution from the interplanetary medium is negligible for all the directions of observations of the present experiment since if current theories concerning the pickup of $He^+$ ions of interstellar origin are correct, the scattering ions are moving with the solar wind speed of $\simeq 400$ km sec$^{-1}$. This corresponds to a Doppler shift of $\simeq 0.4$ Å, which is a factor of 4 higher than the observed half width of the solar line [*Behring et al.*, 1972; *Paresce and Bowyer*, 1973]. The contribution to the observed signal from the ions in the region between the plasmapause and the magnetopause is also well below the threshold of detection in this case, since for the view directions of this flight both the relevant path length and the ion densities involved are extremely small.

This leaves the plasmasphere as the predominant contributor to the observed signal. Because the plasmasphere corotates with the earth and because the solar line width is much larger than the terrestrial scattering width, (1) simplifies to the form

$$I = 10^{-6} \times (\pi e^2/mc) f (\pi F_\nu)_0 p(\theta) N(\mathrm{He}^+) A \qquad (3)$$

In this expression, $(\pi F_\nu)_0$ is the solar line flux at the line center at 1 AU in photons cm$^{-2}$ s$^{-1}$ Hz$^{-1}$, $f$ is the oscillator strength, and $N(\mathrm{He}^+)$ is the illuminated column density of $He^+$ in the plasmasphere. In this study the photoabsorption factor $A$ was computed using the Cira (1965) model 5 and the Chapman grazing incidence function for zenith angles greater than 70°. The Cira (1965) model 5 is appropriate for use here based on the solar 10.7-cm radio flux on the day of the flight. The column density $N(\mathrm{He}^+)$ is computed by integrating the assumed number density profile from the lower to the upper bound of the illuminated column.

The lower bound is assumed to be the intersection of the line of sight with the earth shadow cylinder at 304 Å, whose height is shown in Figure 1*a*. The upper bound is assumed to be the intersection of the line of sight with the plasmapause located at a distance given by the empirical relation [*Binsack*, 1967] $L_p = 6 - 0.6Kp$, where $Kp$ is the average planetary index of geomagnetic activity in the 6-hour period prior to flight.

In the calculations described here we have used the value $\Delta\lambda = 0.15$ Å for the solar line width [*Behring et al.*, 1972] and an integrated solar line flux of $9 \times 10^8$ photons cm$^{-2}$ s$^{-1}$ [*Timothy and Timothy*, 1970]. This last parameter is the appropriate one for use here, according to the solar 10.7-cm flux observed on the day of flight.

We can compute the expected intensity of scattered 304-Å radiation once the appropriate model of the helium ion distribution in the plasmasphere is inserted in (3). The models we employed in our analysis were an isothermal diffusive equilibrium model, a diffusive equilibrium model with a constant positive temperature gradient but without allowance for thermal diffusion, a diffusive equilibrium model with a constant positive temperature gradient with thermal diffusion, a constant density model, and two empirical power law distributions in which the density varies as $R^{-4}$ and $R^{-3}$. Although the density distributions we have assumed in this analysis are strictly valid only along magnetic field lines, the magnetic dip angle at the altitudes and geomagnetic latitudes of the observations is sufficiently high to justify using these distributions as purely vertical profiles.

*Isothermal diffusive equilibrium model.* In this model we assume that the number density of He$^+$ satisfies the relation [*Bauer*, 1969]

$$\partial \ln [n(\mathrm{He}^+)]/\partial z' = -\bar{m}g_0/k(T_e + T_i) \qquad (4)$$

where $z'$ is the geopotential altitude, $\bar{m}$ is the mean ionic mass, $g_0$ is the acceleration of gravity at a convenient reference level, and $T_e$ and $T_i$ are the constant electron and ion temperatures, respectively, in the plasmasphere. Since little is known about the altitude variations of $T_e$ and $T_i$ separately at these altitudes, we have assumed $T_e = T_i = 1000$°K, as suggested by the Cira (1965) model 5 and by theoretical calculations [*Schunk and Walker*, 1969; *Banks*, 1966]. The results obtained using this model to compute the expected 304-Å intensities are shown by the dotted line in Figure 1*c*. In constructing this model we assumed that the value of $n(\mathrm{He}^+)$ at the reference level of 1600 km is $6 \times 10^5$ cm$^{-3}$ and that H$^+$ is the predominant ion ($\bar{m} \simeq m_H$). The predicted spatial variations are clearly inconsistent with the observations. Since this is true for any choice of the variable parameter, we conclude that the He$^+$ ion is not in isothermal diffusive equilibrium in the plasmasphere, confirming our earlier results [*Paresce et al.*, 1973].

*Diffusive equilibrium without thermal diffusion.* In this model we assume He$^+$ to be in diffusive equilibrium in an exosphere where the ion temperature increases with altitude with a constant positive gradient $\beta$, and we neglect the effects of thermal diffusion. Under these conditions the radial dependence of the ionized helium ion density is given by the relation [*Bauer*, 1969]

$$n(\mathrm{He}^+) = n_0(\mathrm{He}^+) \cdot \exp\left\{-\int_{R_0}^{R}\left[\left(m_{\mathrm{He}} - \frac{\bar{m}}{2}\right)\frac{g_0}{kT} + \frac{\beta}{T}\right]\frac{g}{g_0}\,dR\right\} \qquad (5)$$

where $n_0(\mathrm{He}^+)$ is the He$^+$ number density at the reference altitude of 1600 km. We have computed the expected intensities using this distribution with the assumption that H$^+$ is the predominant ion ($\bar{m} \simeq m_H$). We have considered $n_0(\mathrm{He}^+)$ and $\beta$ the variable parameters. The results are very sensitive to the value of the temperature gradient $\beta$. A best fit to the 304-Å data shown in Figure 1 is obtained with $\beta = 0.5$°K km$^{-1}$ and $n_0(\mathrm{He}^+)$ at 1600 km of $2 \times 10^5$ ions cm$^{-3}$ ($\chi^2 = 12.5$, confidence 42%). The results of these calculations using these values of the free parameters are shown in Figure 1*c* by the thick solid line. The value of $\beta$ obtained in this way is in good agreement with theoretical estimates [*Bauer*, 1966] and satellite observations [*Colin et al.*, 1969].

*Diffusive equilibrium with thermal diffusion.* In this model we assume that the exospheric He$^+$ density profile is influenced by thermal diffusion as discussed by *Walker* [1967] and *Schunk and Walker* [1969]. Following these last authors we assume that the ion temperature $T_i$ increases with altitude with a constant positive temperature gradient $\beta$ from a value of $T_i = 1000$°K at an altitude of 600 km. As in the previous model, we have considered $\beta$ and $n_0(\mathrm{He}^+)$, the He$^+$ density at 1600 km, the variable parameters. A best fit to the 304-Å data shown in Figure 1*c* is obtained with the model using $\beta = 1.2$°K km$^{-1}$ and $n_0(\mathrm{He}^+) = 1.5 \times 10^5$ cm$^{-3}$ ($\chi^2 = 23.9$, confidence 4%). The results of these calculations using these values of the variable parameters are shown in Figure 1*c* by the dash-dot line.

*Constant density model.* In this model we assume, following *Meier and Weller* [1972] and *Paresce et al.* [1973], that the He$^+$ density throughout the plasmasphere is independent of altitude. The latitude variation of the ion density was accounted for by an equatorial trough of the type observed by *Chandra et al.* [1970]. The best fit to the observational data can be obtained with this model by using $n(\mathrm{He}^+) = 500 \pm 150$ ions cm$^{-3}$ and assuming a negligible equatorial ion trough. The calculations using this model, however, are very insensitive to this last parameter owing to the large attenuation of the atmosphere above the rocket when the detector's line of sight intersects the equatorial regions in the last part of the flight. For this reason the data cannot be used to yield reliable information on the depth of the equatorial trough. The density obtained this way is compatible with previous 304-Å airglow results. The predictions of such a model are shown in Figure 1*c* by the faint solid line. This model predicts lower intensities at the beginning of the flight and slightly higher intensities toward the end of the flight than the intensities observed, and the overall fit is not good ($\chi^2 = 73.8$, confidence $<1\%$).

*Power law density distributions.* In these models we assume the He$^+$ density in the plasmasphere to vary with radial distance $R$ as

$$n(\mathrm{He}^+) = n_0(\mathrm{He}^+)(R_0/R)^k \qquad (6)$$

with $k = 3$ or 4. The $R^{-4}$ distribution is a rough approximation to the altitude distribution observed by *Chappell et al.* [1970] with a mass spectrometer on Ogo 5, and the $R^{-3}$ model approximates the behavior of a collisionless ion exosphere [*Bauer*, 1969]. The free parameter in these models is the helium ion density at the reference altitude of 1600 km. The best fit models are shown in Figure 1*c* as a long-dashed line ($k = 4$) and a short-dashed line ($k = 3$), corresponding to $n_0(\mathrm{He}^+)$ at 1600 km of $10^4$ and $6 \times 10^3$ cm$^{-3}$,

respectively. A comparison of the predicted intensity variations with the variations of the 304-Å shadow height as a function of rocket time shown in Figure 1*a* shows that the $k = 4$ model overestimates the observed intensities at the lower altitudes and underestimates them at the higher altitudes ($\chi^2 = 142$, confidence $\ll 1\%$). The $R^{-3}$ model, however, is roughly consistent with the observations ($\chi^2 = 25.5$; confidence 2%).

## Discussions and Conclusions

The best fit to the observations reported in this paper is obtained by a model in which plasmaspheric He+ is in diffusive equilibrium in an atmosphere having a constant positive temperature gradient of 0.5°K km$^{-1}$. In this model the ion density initially decreases rapidly with increasing altitude but levels off in the region of 1–2 radii from the earth's surface. Because of this rapid initial decline, the satisfactory fit shown in Figure 1 can be obtained only by assuming that the ion density at 1600 km is $2 \times 10^5$ ions cm$^{-3}$. This value is several orders of magnitude higher than the values commonly observed in situ at these altitudes [*Taylor et al.*, 1965, 1968, 1971; *Brinton et al.*, 1969, 1971; *Chandra et al.*, 1970; *Chappell et al.*, 1970] and is more than an order of magnitude higher than the most optimistic theoretical predictions [*Hanson*, 1962; *Bates and Patterson*, 1962; *Bauer*, 1969].

We cannot entirely rule out the possibility that the He+ density at 1600 km was indeed this high at the time of our flight because of the large differences in the existing in situ satellite data and also because of the large variations that this parameter is known to undergo with such factors as local time, latitude, and geomagnetic conditions. Nonetheless, we feel that this possibility is highly unlikely, and on this basis we reject this model.

In the case of the constant density plasmaspheric model, the predicted intensities and spatial variations do not lead to a satisfactory fit to the observations, but the density derived is more compatible with our present knowledge. Two aspects of this model can be compared with in situ satellite data: the altitude variation and the number density at some reference altitude. The altitude variation is roughly consistent with the observations of *Taylor et al.* [1965]. Unfortunately, it is inconsistent with the measurement of *Chappell et al.* [1970] of the same parameter. The number density required, although it is consistent with theoretical predictions [*Bauer*, 1969], is approximately a factor of 10 or more above that measured by *Taylor et al.* [1965]. We reject this model here on the basis that it cannot reproduce our resonant scattering data. It should be noted that *Meier and Weller* [1972] and *Paresce et al.* [1973] found that their data could be well fit by this type of model.

As we have shown, a diffusive equilibrium model that allows for thermal diffusion and an $R^{-3}$ distribution both fit our resonant scattering data reasonably satisfactorily but an $R^{-4}$ distribution does not. An $R^{-3}$ distribution suggests a collisionless ion exosphere of the type expected beyond the plasmapause [*Bauer*, 1969]. The number densities at 1600 km ($1.5 \times 10^5$ for the thermal diffusion model and $6 \times 10^5$ for the $R^{-3}$ power law model) needed for compatibility with the observational data are well within the acceptable range of He+ densities in the exosphere.

In view of this result we have reexamined our 1970 data [*Paresce et al.*, 1973] for compatibility with an empirical power law and a thermal diffusion distribution. The best fits to these data using these distributions of He+ along a magnetic field line (the 1970 measurements were made at the geomagnetic equator) yield confidence levels substantially less than 1% for the power law models but about 7% for the thermal diffusion model. This result may be taken either as an indication of experimental difficulties or of a real temporal variation of the He+ distribution. Since temporal variations are commonly observed in the electron and proton distribution [*Chappell*, 1972; *Carpenter and Park*, 1973], the latter explanation seems more likely.

We conclude that our data on the nighttime He II 304-Å radiation can be used to exclude an isothermal diffusive equilibrium, a constant density, or an empirical $R^{-4}$ power law distribution of the plasmaspheric He+ distribution. We cannot a priori exclude a diffusive equilibrium model with a constant temperature gradient, but the suggested He+ density at lower altitudes is grossly contradicted by in situ satellite measurements of this parameter. The models that lead to the most satisfactory compromise between reasonable number densities and a satisfactory fit to the observed spatial variations are the thermal diffusion and the empirical $R^{-3}$ power law models. It seems clear that a simultaneous He II 304-Å observation and satellite in situ measurement of the He+ density at any point on the line of sight would place a significant constraint on models of the exospheric He+ distribution.

*Acknowledgments.* We wish to thank Michael Lampton and Ray Cruddace for assistance with technical aspects of the experiment and pay load integration. We also thank Bruce Margon for assistance in data reduction and a referee for helpful suggestions regarding the thermal diffusion model. This work was supported by NASA grant NGR 05-003-450.

* * *

The Editor thanks R. R. Meier and T. Ogawa for their assistance in evaluating this paper.

## References

Banks, P. M., Thermal conduction and ion temperatures in the ionosphere, *Earth Planet. Sci. Lett.*, *1*, 270, 1966.

Bates, D. R., and T. N. L. Patterson, Helium ions in the upper atmosphere, *Planet. Space Sci.*, *9*, 599, 1962.

Bauer, S. J., Hydrogen and helium ions, *Ann. Geophys.*, *22*, 247, 1966.

Bauer, S. J., Diffusive equilibrium in the topside ionosphere, *Proc. IEEE*, *57*, 1114, 1969.

Behring, W. E., L. Cohen, and U. Feldman, The solar spectrum: Wavelengths and identification from 60-385 A, *Astrophys. J.*, *175*, 493, 1972.

Binsack, J. H., Plasmapause observations with the MIT experiment on Imp 2, *J. Geophys. Res.*, *72*, 5231, 1967.

Bowyer, C. S., F. Paresce, M. Lampton, and J. Mack, A far ultraviolet photometer for space research, *Planet. Space Sci.*, *18*, 835, 1970.

Brandt, J. C., and J. W. Chamberlain, Interplanetary gas, 1, Hydrogen radiation in the night sky, *Astrophys. J.*, *130*, 670, 1959.

Brinton, H. C., R. A. Pickett, and H. A. Taylor, Jr., Diurnal and seasonal variation of atmospheric ion composition; correlation with solar zenith angle, *J. Geophys. Res.*, *74*, 4064, 1969.

Brinton, H. C., J. M. Grebowsky, and H. G. Mayr, Altitude variation of ion composition in the mid-latitude trough region: Evidence for upward plasma flow, *J. Geophys. Res.*, *76*, 3738, 1971.

Carlson, R. W., Possibility of O III λ304-Å emissions in the extreme ultraviolet airglow, *J. Geophys. Res.*, *77*, 6282, 1972.

Carpenter, D. L., and C. G. Park, On what ionospheric

workers should know about the plasmapause-plasmasphere, *Rev. Geophys. Space Phys., 11,* 133, 1973.

Chandra, S., B. E. Troy, J. L. Donley, and R. E. Bourdeau, Ogo 4 observations of ion composition and temperatures in the topside ionosphere, *J. Geophys. Res., 75,* 3867, 1970.

Chappell, C. R., Recent satellite measurements of the morphology and dynamics of the plasmasphere, *Rev. Geophys. Space Phys., 10,* 951, 1972.

Chappell, C. R., K. K. Harris, and G. W. Sharp, A study of the influence of magnetic activity on the location of the plasmapause as measured by Ogo 5, *J. Geophys. Res., 75,* 50, 1970.

Colin, L., S. W. Dufour, and D. S. Willoughby, An interpretation of Ogo light ion abundance measurements, *J. Geophys. Res., 74,* 1863, 1969.

Hanson, W. B., Upper atmosphere helium ions, *J. Geophys. Res., 67,* 183, 1962.

Hoffmann, T. H., Studies of the composition of the ionosphere with a magnetic deflection mass spectrometer, *Int. J. Mass. Spectrom. Ion Phys., 4,* 835, 1970.

Meier, R. R., and C. S. Weller, EUV resonance radiation from helium atoms and ions in the geocorona, *J. Geophys: Res., 77,* 1190, 1972.

Ogawa, T., and T. Tohmatsu, Sounding rocket observation of helium 304-Å and 584-Å glow, *J. Geophys. Res., 76,* 6136, 1971.

Paresce, F., and S. Bowyer, On the backscatter of solar He II 304 Å radiation from interplanetary $He^+$, *Planet. Space Sci., 21,* 1795, 1973.

Paresce, F., S. Bowyer, and S. Kumar, Observations of the He II 304-Å radiation in the night sky, *J. Geophys. Res., 78,* 71, 1973.

Schunk, R., and J. C. G. Walker, Thermal diffusion in the topside ionosphere for mixtures which include multiply charged ions, *Planet. Space Sci., 17,* 853, 1969.

Taylor, H. A., H. C. Brinton, and C. R. Smith, Positive ion composition in the magnetoionosphere obtained from the Ogo A satellite, *J. Geophys. Res., 70,* 5769, 1965.

Taylor, H. A., H. C. Brinton, M. W. Pharo, and N. K. Rahman, Thermal ions in the exosphere: Evidence of solar and geomagnetic control, *J. Geophys. Res., 73,* 5521, 1968.

Taylor, H. A., J. M. Grebowski, and W. J. Walsh, Structured variations of the plasmapause: Evidence of a corotating plasma tail, *J. Geophys. Res., 76,* 6806, 1971.

Timothy, A. F., and J. G. Timothy, Long-term intensity variations in the solar helium II Lyman $\alpha$ line, *J. Geophys. Res., 75,* 6950, 1970.

Walker, J. C. G., Thermal diffusion in the topside ionosphere, *Planet. Space Sci., 15,* 1151, 1967.

Young, J. M., C. S. Weller, C. Y. Johnson, and J. C. Holmes, Rocket observations of the far UV nightglow at Lyman $\alpha$ and shorter wavelengths, *J. Geophys. Res., 76,* 3710, 1971.

(Received May 16, 1973;
accepted September 19, 1973.)

# OBSERVATIONS OF He I 584 Å NIGHTTIME RADIATION: EVIDENCE FOR AN INTERSTELLAR SOURCE OF NEUTRAL HELIUM

A74-31956

FRANCESCO PARESCE, STUART BOWYER,
AND SHAILENDRA KUMAR

Reprinted for private circulation from
THE ASTROPHYSICAL JOURNAL, Vol. 187, No. 3, Part 1, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved.
PRINTED IN U.S.A.

THE ASTROPHYSICAL JOURNAL, 187:633–639, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

# OBSERVATIONS OF He I 584 Å NIGHTTIME RADIATION; EVIDENCE FOR AN INTERSTELLAR SOURCE OF NEUTRAL HELIUM

FRANCESCO PARESCE, STUART BOWYER, AND SHAILENDRA KUMAR*
Department of Astronomy and Space Sciences Laboratory, University of California, Berkeley
*Received 1973 April 9; revised 1973 August 17*

## ABSTRACT

The intensity and spatial variations of the He I 584 Å radiation in the night sky were measured to an altitude of 264 km from a sounding rocket launched from Thumba, India. The flux varied in intensity with look angle from a high of 16 ± 5 rayleighs to a low of < 1 rayleigh. The data obtained are presented in the form of an all-sky map. Since the peak flux levels observed are approximately two orders of magnitude larger than that expected from geocoronal resonant scattering, we have compared our results with those expected from two interplanetary sources of neutral helium: the penetration of the local interstellar medium into the solar neighborhood and the dust de-ionization of solar-wind alpha particles. The dust de-ionization-model predictions using currently accepted input parameters appear to be inconsistent with the observations. The interstellar helium source is compatible with the data to within the uncertainties of the models.

*Subject headings:* airglow — interplanetary medium — solar wind — spectra — ultraviolet

## I. INTRODUCTION

On 1970 March 10, we flew three extreme-ultraviolet photometers to an altitude of 264 km on a Nike Tomahawk rocket. We have previously reported results obtained with two of these detectors on the intensity and spatial variations of the hydrogen Lα and He II 304 Å glow (Paresce, Kumar, and Bowyer 1972; Paresce, Bowyer, and Kumar 1973). These fluxes are believed to be geophysical in origin. We report here observations made with the third of these instruments on the spatial distribution and intensity of the He I 584 Å nightglow. We believe these measurements to be the first observations of this radiation. We advance the possibility that a major portion of the He I 584 Å nightglow has an astrophysical rather than a geophysical origin.

## II. INSTRUMENTATION AND FLIGHT

The measurements were made using a 1-cm diameter channel electron multiplier detector with a thin metallic film of approximately 1000 Å of aluminum which was employed as a bandpass filter. The combination of photocathode sensitivity and filter transmission restricted the bandpass of this instrument to the approximate range of 170–800 Å. Instrumentation of this type has been described in detail by Bowyer *et al.* (1970).

Prior to flight, an absolute calibration of the detector was carried out at several wavelengths, including He II 304 Å, He I 584 Å, and H I 1216 Å lines with the use of a high-intensity continuous-discharge hollow-cathode extreme-ultraviolet source (Paresce, Kumar, and Bowyer 1971) and a Seya-Namioka monochromator. The quantum efficiency of the detector was determined by reference to gold, tantalum, and platinum photodiodes, which were employed as absolute intensity standards. Measurements of the photoelectric yield of these metals in the extreme-ultraviolet have been summarized by Cairns and Samson (1966). The values reported by these authors agree to an accuracy of ±30 percent.

Circular collimation of 10° FWHM was provided by a drilled-hole metal collimator. The angular response of the collimator and the effective area of the detector were measured by using a highly collimated beam of 584 Å radiation. Because substantial fluxes of electrons and protons with energies in the 0.1–10 keV range have been detected on previous rocket flights (Hill *et al.* 1970) and satellite measurements at low altitudes (Heikkila 1971), the collimator was combined with a magnetic broom for charged-particle rejection. The charged-particle sensitivity of the complete instrument was checked with an intense beam ($\simeq 10^8$ particles $s^{-1}$ $cm^{-2}$) of 0–70 keV electrons and protons generated by an electrostatic accelerator. The photometer was completely insensitive to electrons with energies less than 30 keV and protons with energies less than 6 keV. Table 1 summarizes the relevant characteristics of the photometer.

The rocket was launched from Thumba, India (geomagnetic latitude 0°) on 1970 March 10, at 2:05 IST (India Standard Time). At the time of flight the Sun was in the east, 60° below the horizon. The detector was mounted looking out the side of the rocket with a view perpendicular to the spin axis. During flight the rocket followed a stable, torque-free precessional motion. The rocket aspect as a function

* Present address: Kitt Peak National Observatory, Tucson, Arizona, 85717.

TABLE 1

PHOTOMETER CHARACTERISTICS

| Parameter | Value |
|---|---|
| Bandpass (to 10 percent transmission) | 170–800 Å |
| Effective area | 0.75 $cm^2$ |
| Effective solid angle | 0.015 sterad |
| Quantum efficiency (counts/photon): | |
| 304 Å | $3.6 \times 10^{-2}$ |
| 584 Å | $1.3 \times 10^{-2}$ |
| 1026 Å | $2.7 \times 10^{-6}$ |
| 1216 Å | $3.8 \times 10^{-7}$ |

of time was derived from an analysis of onboard magnetometer signals in combination with a variety of flight-detector readings. The hydrogen Lα detector and two detectors monitoring the oxygen lines at 5577 and 6300 Å were especially useful in this regard. The aspect solution derived in this manner was good to within ±10°. During flight the rocket had a spin rate of 5 rps. It precessed with a coning angle of ≃40° and a period of 30 s about an axis directed near the zenith. As a result of this motion, the detector repeatedly scanned a portion of sky between zenith angles of approximately 50° and 130°.

## III. RESULTS

The detector was turned on at 92 km and a small but relatively constant signal was recorded up to 150 km. Above 150 km the signal increased rapidly with altitude. The reverse of the behavior was repeated on the downleg of the flight. The constant count rate of ≃8 counts $s^{-1}$ observed in the 100–150-km region was interpreted as being due to hydrogen Lα flux. The Lα intensity necessary to produce the observed signal is consistent with measurements of this line made during the flight (Paresce *et al.* 1972). Because in the altitude range of our measurements Lα is expected to remain fairly constant (Donahue 1966), we have subtracted this background from all the data. In figure 1*a* we show the background-corrected count rate from the instrument as a function of altitude for the particular view direction of 0° azimuth (east), 70° zenith angle. Each data point corresponds to an average of 12 scans over this particular area of the sky.

The data presented in figure 1*a* may be converted into emission rates once the radiation detected is identified. Strong extreme-ultraviolet airglow lines which are expected in the upper atmosphere at night and which should be roughly compatible with the observed altitude profile are the O II 834 Å, He I 584 Å, and He II 304 Å lines. We believe that the flux detected is not O II 834 Å radiation since the low quantum efficiency of the photometer to this radiation would require an input intensity of this radiation of at least 0.5 kilorayleighs. This intensity is approximately a factor of 20 higher than what is expected on the basis of resonance scattering of the solar 834 Å line by $O^+$

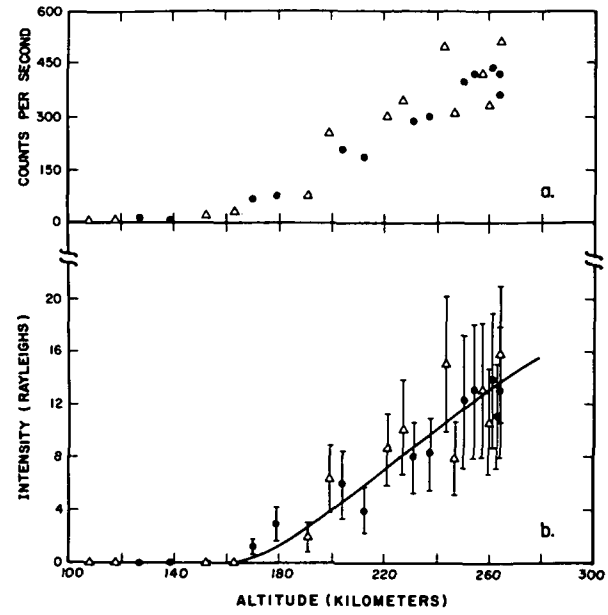



FIG. 1.—(*a*) Count rate of the aluminum filter detector as a function of rocket altitude. (*b*) Intensity in rayleighs of the He I 584 Å line as a function of rocket altitude. The solid line represents the theoretical extinction curve for an overhead source of He I, 584 Å radiation.

in the plasmasphere in the amounts measured from OGO-5 (Chappell, Harris, and Sharp 1970).

Another detector, similar to the device described here but with a nominal 1000 Å aluminum and 270 Å carbon filter, which is sensitive mainly to He II 304 Å radiation, was included in the flight package to permit us to distinguish between the two helium lines. The He II data observed have been analyzed (Paresce *et al.* 1973), and the results obtained were used to assess the contribution of 304 Å radiation to the data shown in figure 1*a*. In figure 1*b* we show the intensity of the 584 Å radiation based on the assumption that the data displayed in figure 1*a* are an admixture of He II 304 Å and He I 584 Å radiation, and that we have correctly accounted for the He II contribution to the observed signal. The error bars in this figure reflect the statistical fluctuations in the measurements and reading errors. Any error in the assumed photoelectric yields of the calibration standards would have the effect of changing the ordinate in this figure by a constant scale factor.

We have carried out a partial check of our method of data analysis by fitting the data in figure 1*b* to a theoretical extinction curve for He II 584 Å radiation. This curve is shown as a solid line in figure 1*b* and was obtained by assuming an overhead source of radiation with absorption due to photoionization of $O_2$, $N_2$, and O above the rocket. This approximation gives absorption profiles which are within ±20 percent of those obtained by detailed model atmosphere calculations (Kumar, Bowyer, and Lampton 1973). The number

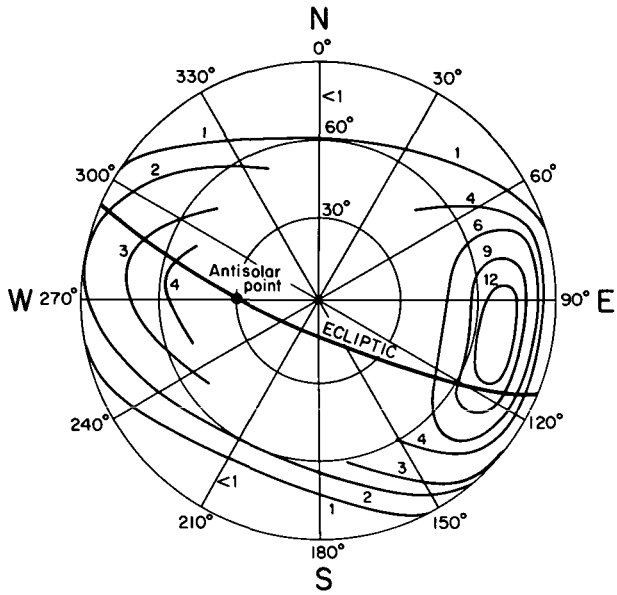



FIG. 2.—Isophotes in rayleighs of the observed He I 584 Å radiation plotted on a zenith-centered upper hemisphere of the night sky.

densities of the absorbing constituents were obtained from the CIRA (1965) Model 5. This model was appropriate for use here based on the solar 10.7-cm radio flux on the day of our flight. The photoionization cross-sections were taken from Hall, Schweizer, and Hinteregger (1965). The fit is satisfactory ($\chi^2 = 32$; confidence level is 6 percent). In the following we proceed on the assumption that our method of analysis is correct and the radiation discussed here is, in fact, He I 584 Å radiation.

In a given look direction the intensity of the radiation changes very little with altitude above 250 km. A total of 113 s was spent above this altitude observing the directional variations of this flux. In figure 2 we present the 584 Å data taken above 250 km as isophotes on a local zenith-centered all-sky map. The Earth's equator and the ecliptic plane are displayed on the diagram together with the position of the antisolar direction. The isophotes in this case seem to be centered slightly away from the solar azimuth in the direction of the ecliptic. The peak intensity of 12 ± 4 rayleighs was observed about 30° above the horizon in an easterly direction. The isophotes show a strong asymmetry with most of the emission concentrated near the eastern and western horizons. Very little radiation is observed to originate in the northern and southern skies. A measurable flux of 584 Å radiation (2–3 rayleighs) is found to originate from directions 40° below the eastern and western horizons.

The results reported here are inconsistent with the work of Young *et al.* (1971), who obtained an upper limit to the He I 584 Å nightglow which was considerably below the flux levels we observed. The reason for this discrepancy is not clear. Riegler and Garmire (1972) have made a preliminary report of observations of a nighttime 584 Å flux, and the flux levels they observed were similar to those reported here.

## IV. INTERPRETATION

We have considered a number of possibilities as to the source of the signal we observed. At the altitudes of our observations the helium geocorona is optically thick to 584 Å radiation. Meier and Weller (1972) have compared a preliminary report of our data (Paresce *et al.* 1970) with the predictions of a geocoronal resonance scattering model. These authors found our results to be about two orders of magnitude greater than what could be expected assuming a model consistent with available He I 584 Å dayglow observations.

A second possible geophysical source of counts in our detector may be high energy ($E > 6$ keV) atmospheric protons. Substantial fluxes of atmospheric charged particles are known to occur sporadically on the equator in the evening hours (Heikkila 1971) and their presence is strongly correlated with the occurrence of spread $F$ (Heikkila and Winningham 1973). Protons of this energy would exhibit a flux-versus-altitude profile very similar to the one expected for 584 Å radiation and shown by the data in figure 1. To account for the approximately 500 counts $s^{-1}$ registered by our photometer we would need $\simeq 5 \times 10^4$–$10^5$ protons $cm^{-2}$ $s^{-1}$ $sterad^{-1}$ at energies above 7 keV, the energy at which protons reach 50 percent transmission through the aluminum filter. A flux of this magnitude is not ruled out by Heikkila's (1971) data since the observed energy spectrum may be broad and extend up to a few kiloelectron volts. Because of its highly peaked pitch angle distribution (Heikkila 1971) the proton flux would be expected to decrease when observing in the north and south in a manner reminiscent of the data presented in figure 2.

We argue against this interpretation, however, since the marked east-west asymmetry in our data contrast with Heikkila's (1971) observations of an absence of such an asymmetry in the higher energy protons. In addition, ionograms during the rocket flight showed no evidence of strong irregularities of the type usually associated with spread $F$.

Detectors identical to the one described in this paper were flown in 1972 and 1973 on two sounding rockets launched from White Sands Missile Range. Detailed results from these flights will be given in a forthcoming paper. A preliminary analysis of the data indicates that the signals registered during these subsequent flights are of the same magnitude (a few hundred counts $s^{-1}$) as the ones observed in the Thumba 1970 flight. Heikkila's (1971) observation of energetic charged particles shows that the particles he observed are restricted to a small region around the

geomagnetic equator and the auroral regions. No evidence of their presence outside of these regions has been found. For these reasons we believe it is unlikely that the observed signals in the Thumba 1970 flight were due to energetic protons.

Another possibility we have investigated to account for our observations is that of an extraterrestrial 584 Å radiation source. In a separate paper (Paresce and Bowyer 1973) we have investigated the He I 584 Å radiation field expected outside the Earth's helium geocorona due to two interplanetary sources of neutral helium: the penetration of the local interstellar medium into the solar neighborhood and the dust de-ionization of solar wind alpha particles. Radiation from either of these sources would not be absorbed by helium in the Earth's geocorona if the interplanetary 584 Å radiation were Doppler shifted from the absorption core of the terrestrial helium by only 8 mÅ. This shift corresponds to a relative velocity of the scattering medium and the Earth of $\simeq 4$ km s$^{-1}$.

In the following we compare our data with theory by generating the radiation field for each of these sources for the geometry applicable at the time of our flight and then correcting for absorption within the Earth's geocorona. The procedure employed to generate the radiation field has been described in detail elsewhere (Paresce and Bowyer 1973). In brief, if the observer's line of sight makes an angle $\lambda$ with the Sun/Earth radius vector, and an angle $\beta$ with the ecliptic plane, the intensity of the scattered 584 Å radiation in photons s$^{-1}$ cm$^{-2}$ sterad$^{-1}$ at a frequency $\nu$ is assumed to be given by

$$I_\nu(\lambda, \beta) = \frac{1}{4\pi} \int p(\theta) n(r) F_\nu(r) a_\nu \, dr \, . \qquad (1)$$

In this formula $n(r)$ represents the helium number density as a function of distance $r$ from the Sun. The solar 584 Å flux $F_\nu$ is represented by a simple Gaussian which falls off in intensity as the square of the distance from the Sun, and the phase function for scattering, $p(\theta)$, is a function of the angle $\theta$ between the Sun, the scattering particle, and the observer. The scattering coefficient of the atoms is a Gaussian with the central frequency Doppler shifted by the radial velocity of the scatterer with respect to the Sun and with a width which is dependent upon the temperature of the gas.

For neutral helium due to the penetration of the local interstellar medium into the solar system we have inserted in equation (1) the helium densities derived by Holzer and Axford (1971) for the case $\mu = 0$ (radiation pressure negligible) and the photoionization rate $\beta_0 = 0.62 \times 10^{-7}$ s$^{-1}$. In the wake regions, the formulae developed by Feldman, Lange, and Scherb (1972) specifically for these regions were used. The interstellar wind has been assumed to approach the Sun from the direction of ecliptic longitude 270°, ecliptic latitude 0° (Thomas 1972). The helium number density in the local interstellar medium $N$(He), the temperature $T$ and velocity of approach $V$ of the interstellar medium, and the solar line width in angstroms, $\Delta\lambda_s$, are variable parameters.

For the dust de-ionization model we have assumed that the steady-state helium density is given by

$$n(r, z) = \left[\frac{\phi_0}{I_0} G\left(\frac{r_0}{r}\right) e^{-Kz/r_0}\right], \qquad (2)$$

where the radial distribution of the helium density has been taken from Banks (1971). In this formalism $\phi_0$ is the solar wind flux of $He^{++}$, $I_0$ is the helium photoionization rate at 1 a.u., and $G$ is the dust particle geometrical factor. For the helium distribution above the ecliptic plane we use the form suggested by Aller *et al.* (1967) for the distribution of dust forming the zodiacal light with $z$ representing the height above the plane of the ecliptic in a.u. and $1/K$ the neutral helium scale height. Banks (1971) has shown that once helium atoms are produced by the de-ionization process they follow direct Keplerian orbits around the Sun. We have made the simplifying assumption here that particles above the ecliptic plane have trajectories which are in a plane parallel to the ecliptic plane. The intensity emitted at any given point was Doppler shifted by $(1 \pm v/c)\nu_0$ with $v$ the relative velocity between the atoms in the emitting region and the Earth. The resulting emission profile was then integrated over frequency, and a region 8 mÅ wide on both sides of the rest frequency was removed to account for absorption by the terrestrial helium.

The integrals in equation (1) were carried out using standard numerical techniques. For the interstellar backscatter models the calculations were extended out to a few hundred a.u.'s where the contribution to the total intensity became negligible, and for the dust models the integration was performed out to 10 a.u. where the dust density becomes negligible (Southworth 1967).

To facilitate the comparison between the calculated flux and the observational data, the data were corrected for residual absorption due to atmospheric $O_2$, $N_2$, and O in the line of sight and were then replotted in ecliptic coordinates. These corrections were made by using number densities from CIRA (1965) Model 5 and the Chapman grazing-incidence function for zenith angles greater than 70°. The results are shown in figure 3*a*. In this figure the ecliptic longitude $\lambda$ is measured westward from the vernal equinox.

In figure 3*b* we display the results for the interstellar helium model with $N$(He) = 0.008 cm$^{-3}$, $T = 10^2$ °K, $V = 10$ km s$^{-1}$, and $\sqrt{\pi}\lambda_s = 0.027$ Å. The helium number density in the local interstellar medium was obtained by combining the hydrogen number densities obtained from recent extraterrestrial L$\alpha$ scattering data (Thomas 1972), with a hydrogen-to-helium ratio derived from measurements of the

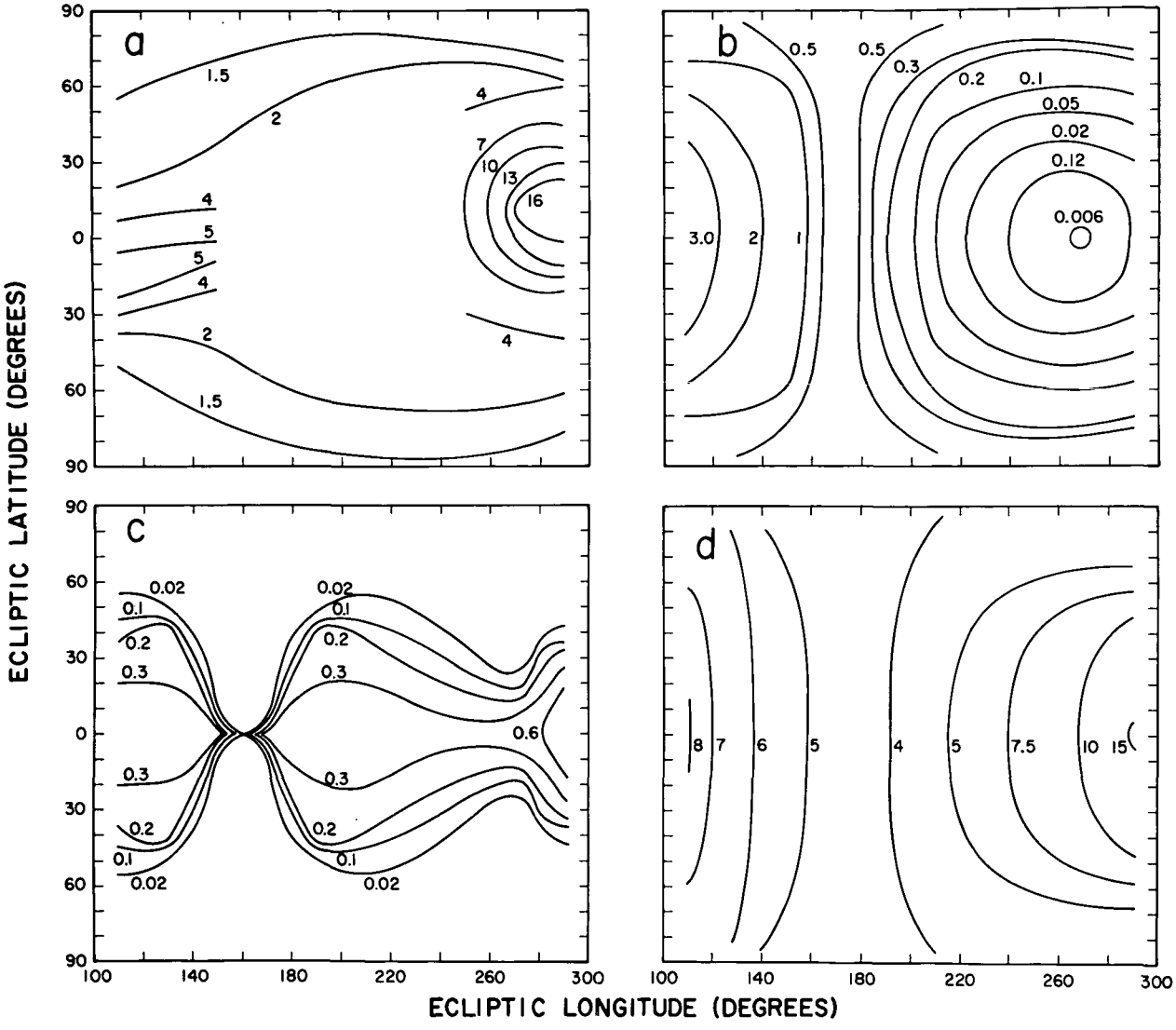



FIG. 3.—Isophotes in rayleighs of the interplanetary helium glow at 584 Å. (*a*) observed and corrected for pure absorption, plotted on an ecliptic-centered map of the night sky. (*b*) computed with the interstellar helium model for $N(\mathrm{He}) = 0.008\ \mathrm{cm}^{-3}$, $T = 10^{2.0}$° K, $V = 10\ \mathrm{km\ s}^{-1}$, and $\sqrt{\pi}\Delta\lambda_s = 0.027$ Å. (*c*) computed with the dust de-ionization model for $K = 5.0$, $G = 5 \times 10^{-18}$ $\mathrm{cm}^{-1}$, $\phi_0 = 8 \times 10^6\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$, $\sqrt{\pi}\Delta\lambda_s = 0.027$ Å. (*d*) with the interstellar helium model for $N(\mathrm{He}) = 0.04\ \mathrm{cm}^{-3}$, $T = 10^{4.0}$° K, $V = 5\ \mathrm{km\ s}^{-1}$, and $\Delta\lambda_s = 0.15$ Å.

strength of radio-frequency transitions of recombining helium ions in the interstellar medium (Palmer *et al.* 1969). The temperature and speed of approach of the local interstellar medium was taken from the hydrogen Lα data (Thomas 1972). The helium solar line width was taken from airglow observations of the daytime He I 584 Å glow (Donahue and Kumer 1971; Kumar *et al.* 1973).

The expected intensities from the dust de-ionization model are shown in figure 3*c* with $K = 5.0$, $G = 5 \times 10^{-18}$ $\mathrm{cm}^{-1}$, $\phi_0 = 8 \times 10^6$ $\mathrm{cm}^{-2}$ $\mathrm{s}^{-1}$ and $\sqrt{\pi}\Delta\lambda_s = 0.027$ Å. The value of $K$ was chosen on the basis of zodiacal-light observations (Aller *et al.* 1967), and $\phi_0$ comes from Bame *et al.*'s (1968) observations of the solar wind. The value of the dust geometrical factor is quite uncertain. The relevant data available in 1970 have been summarized by Banks (1971); these data suggest that this parameter falls within the range of $10^{-16}$–$10^{-21}$ $\mathrm{cm}^{-1}$. Recent satellite measurements of resonantly scattered Lα radiation can be used to set upper limits on this parameter assuming the de-ionization process is reasonably well understood.

Employing the Lα data of Bertaux and Blamont (1971) and Thomas and Krassa (1971) with the de-ionization process as suggested by Banks (1971), we find that $G \leqslant 5 \times 10^{-18}$ cm$^{-1}$. In this study we have used this upper limit for our value for $G$; this parameter enters linearly into the intensity calculations and our results may be scaled accordingly.

The interstellar backscatter model and the dust de-ionization model using these input parameters are clearly unable to reproduce the intensity and spatial variations exhibited by the data. The solar 584 Å line profile may not be the simple Gaussian profile we have assumed in the preceding calculations but may exhibit the characteristic double Gaussian profile of the Lα and Lβ lines (Johnson 1972) because the solar chromosphere is moderately thick to this radiation. We have computed the expected 584 Å emission from interstellar helium in this case assuming the width of the central absorption core to be one-half the assumed width of the line as in the case of Lα. The effect of introducing this refinement into our calculations is to eliminate the deep minimum exhibited by the intensity contours in figure 3*b* at an ecliptic latitude of 270°. The overall intensities, however, still remain approximately an order of magnitude below the observations. With larger solar widths the effect of an absorption core is negligible.

In figure 3*d* we show the results of the best-fit interstellar-medium model in which $V = 5$ km s$^{-1}$, $\Delta\lambda_s = 0.15$ Å, $T = 10^4$ ° K, and $N(\mathrm{He}) = 0.04$ cm$^{-3}$. Although these values for these parameters are not the most popular current choices, they do fall within the range of the observational uncertainties. There are still some differences between this model and the data, but these could easily be ascribed to observational uncertainties.

## V. SUMMARY

We have observed an altitude- and aspect-dependent signal which we interpret to be due to He I 584 Å radiation. Since the intensity and spatial variations of 584 Å radiation above 250 km are incompatible with a geocoronal origin, we have investigated the possibility that the observed radiation is emitted by neutral helium in the interplanetary medium. Two sources of neutral helium were examined: interstellar helium penetrating into the solar system, and the de-ionization of solar wind alpha particles on interplanetary dust grains. Only the first source was found to generate sufficient radiation to account for the gross features of the observed emission pattern.

The interstellar helium model fitted the observational data only with a rather large solar line width and a rather high interstellar helium density. The dust de-ionization model appears to be incompatible with the observations when using nominal values of the input parameters but these parameters are only poorly defined at present.

It is reasonable to expect that more extensive observations of the nighttime He I 584 Å radiation will provide information on the temperature and density of the local interstellar medium, and in some directions possibly even on the distribution of dust within the solar system.

We wish to thank Dr. Arthur Aikin and co-workers at the Goddard Space Flight Center for providing space in their payload for this instrumentation and for technical assistance in carrying out the experiment. We thank Dr. Michael Lampton for technical advice on the instrumentation. This research was supported by NASA grant NGR 05-003-450.

## REFERENCES

Aller, L. H., Duffner, G., Dworetsky, M., Gudehus, D., Kilston, S., Leckrone, D., Montgomery, J., Oliver, J., and Zimmerman, E. 1967, in *Zodiacal Light and the Interplanetary Medium*, NASA SP-150, ed. J. L. Weinberg (Washington: U.S. Government Printing Office), p. 243.
Bame, S. J., Hundhausen, A. J., Asbridge, J. R., and Strong, I. B. 1968, *Phys. Rev. Letters*, **20**, 393.
Banks, P. M. 1971, *J. Geophys. Res.*, **76**, 4341.
Bertaux, J. L., and Blamont, J. *Astr. and Ap.*, **11**, 200.
Bowyer, C. S., Paresce, F., Lampton, M., and Mack, J. 1970, *Planet. and Space Sci.*, **18**, 835.
Cairns, R. B., and Samson, J. A. R. 1966, *J. Opt. Soc. Amer.*, **56**, 1568.
Chappell, C. R., Harris, K. K., and Sharp, G. W. 1970, *J. Geophys. Res.*, **75**, 50.
CIRA, 1965. *COSPAR International Reference Atmosphere*, compiled by COSPAR Working Group 4 (Amsterdam: North Holland).
Donahue, T. M. 1966, *Ann. Geophys.*, **22**, 175.
Donahue, T. M., and Kumer, J. B. 1971, *J. Geophys. Res.*, **76**, 145.
Feldman, W. C., Lange, J. J., and Scherb, F. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA SP-308 (Washington: U.S. Government Printing Office), p. 684.
Hall, L. A., Schweizer, W., and Hinteregger, H. E. 1970, *J. Geophys. Res.*, **70**, 105.
Heikkila, W. J. 1971, *J. Geophys. Res.*, **76**, 1076.
Heikkila, W. I., and Winningham, J. D. 1973, Paper given at the 1st Lloyd V. Berkner Symposium on Irregularities in the Equatorial Ionosphere, Dallas, Texas, 1973 May.
Hill, R. W., Grader, R. J., Seward, F. D., and Stoering, J. P. 1970, *J. Geophys. Res.*, **75**, 7267.
Holzer, T. E., and Axford, W. I. 1971, *J. Geophys. Res.*, **76**, 6965.
Johnson, H. E. 1972. *Planet. and Space Sci.*, **20**, 829.
Kumar, S., Bowyer, C. S., and Lampton, M. 1973, *J. Geophys. Res.*, **78**, 1107.
Meier, R. R., and Weller, C. S. 1972, *J. Geophys. Res.*, **77**, 1190.
Palmer, P., Zuckerman, B., Penfield, M., Lilley, A. E., and Mezger, P. G. 1969, *Ap. J.*, **156**, 887.
Paresce, F., and Bowyer, C. S. 1973. *Astr. and Ap.*, **27**, 399.
Paresce, F., Bowyer, S., and Kumar, S. 1973, *J. Geophys. Res.*, **78**, 71.
Paresce, F., Bowyer, C. S., Kumar, S., and Lampton, M. 1970, *EOS Trans. AGU*, **11**, 795.
Paresce, F., Kumar, S., and Bowyer, S. 1971, *Appl. Opt.*, **10**, 1904.
———, 1972, *Planet. and Space Sci.*, **20**, 297.
Riegler, G. R., and Garmire, G. P. 1972. *Bull. A.A.S.*, **4**, 271.

Southworth, R. B. 1967, in *Zodiacal Light and the Interplanetary Medium*, ed. J. L. Weinberg, NASA SP-150, (Washington: U.S. Government Printing Office), p. 179.
Thomas, G. E. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA SP-308 (Washington: U.S. Government Printing Office).
Thomas, G. E., and Krassa, R. F. 1971, *Astr. and Ap.*, **11**, 218.
Young, J. M., Weller, C. S., Johnson, C. Y., and Holmes, J. C. 1971, *J. Geophys. Res.*, **76**, 3710.

D3

# ON THE OPACITY OF THE INTERSTELLAR MEDIUM TO ULTRASOFT X-RAYS AND EXTREME-ULTRAVIOLET RADIATION

RAYMOND CRUDDACE, FRANCESCO PARESCE,
STUART BOWYER, AND MICHAEL LAMPTON

Reprinted for private circulation from
THE ASTROPHYSICAL JOURNAL, Vol. 187, No. 3, Part 1, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved.
PRINTED IN U.S.A.

REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

THE ASTROPHYSICAL JOURNAL, 187:497–504, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

# ON THE OPACITY OF THE INTERSTELLAR MEDIUM TO ULTRASOFT X-RAYS AND EXTREME-ULTRAVIOLET RADIATION

RAYMOND CRUDDACE, FRANCESCO PARESCE,
STUART BOWYER, AND MICHAEL LAMPTON
Astronomy Department and Space Sciences Laboratory, University of California, Berkeley
*Received 1973 July 16; revised 1973 September 4*

## ABSTRACT

The opacity of the interstellar medium at soft X-ray and extreme-ultraviolet wavelengths has been reexamined in the light of recent evidence that the density of interstellar hydrogen in the solar neighborhood may be quite low. It is shown that the distance at which significant attenuation occurs is not negligible, especially in the 50–300 Å region.

*Subject headings:* interstellar matter — opacities — ultraviolet — X-rays

In dim eclipse disastrous twilight sheds
On half the nations, and with fear of change
Perplexes monarchs.

[Milton, *Paradise Lost*]

## I. INTRODUCTION

More than ten years ago Aller (1959) argued that the interstellar medium should be opaque to radiation at wavelengths between 912 Å and the X-ray region of the spectrum. Since then, although considerable work has been carried out to calculate the absorption expected at X-ray wavelengths (Felten and Gould 1966; Bell and Kingston 1967*a*; Brown and Gould 1970) and much has been learned about the structure of the interstellar medium, there has been little inclination to extend these calculations to wavelengths beyond 100 Å, much less to question Aller's basic conclusion.

We have reexamined in detail the transmission of the interstellar medium to soft X-rays and extreme-ultraviolet radiation, using the most recent determinations of the relevant cross-sections and abundances. We conclude that if the density of the interstellar medium is as low and nonuniform as some observations suggest, then soft X-ray or extreme ultraviolet radiation may be observed over considerable distances.

## II. THE EFFECTIVE ABSORPTION CROSS-SECTION

The opacity of the interstellar medium depends on two factors, the absorption cross-section per atom of each species of atom in the medium and the total number of atoms in the line of sight. We have computed an effective absorption cross-section per hydrogen atom as defined by the relation

$$\sigma_e = \frac{1}{n_{\rm H}} \sum n_i \sigma_i ,$$

where $n_{\rm H}$ is the number density of hydrogen atoms and $n_i$ and $\sigma_i$ are, respectively, the number density and cross-section of the species $i$.

Figure 1 shows the photoabsorption cross-section of all elements which contribute appreciably to the opacity between 2000 and 1 Å. The cross-section of the hydrogen atom has been obtained by blending the results of calculations below the Lyman edge by Samson (1966), who used Gaunt factors given by Karzas and Latter (1961), with the results of calculations made at wavelengths below 128 Å (0.1 keV) by Brown and Gould (1970). These calculations agree with the measurements of Beynon and Cairns (1965) at 850 Å, and Denne (1970) between 82 and 23 Å, to within 15 percent. The cross-section of the helium atom has been obtained by matching the work of Bell and Kingston (1967*b*) at longer wavelengths with the results of Brown and Gould (1970) at shorter wavelengths. These calculations agree to within 15 percent with measurements by Samson (1966) and Lowry, Tomboulian, and Ederer (1965) between 500 and 200 Å, Lukirskii, Brytov, and Zimkina (1964) between 230 and 44 Å, and by Denne (1970) between 82 and 44 Å. The cross-sections of the heavier elements were obtained from the sources summarized in table 1.

The adopted relative abundances of the elements are listed in table 2. The helium abundance was obtained from the work of Palmer *et al.* (1969), who measured the strength of radio frequency transitions of recombining helium atoms. The abundances of the heavier elements are those given by Felten and Gould (1966) and by Bell and Kingston (1967*a*), except in the case of neon, for which the reduced value suggested recently by Brown and Gould (1970) was taken. The abundance of aluminum is that quoted by Allen (1963).

The resultant effective cross-section of the interstellar medium, $\sigma_e$, has been plotted against wavelength in figure 2, under the assumption that ionization

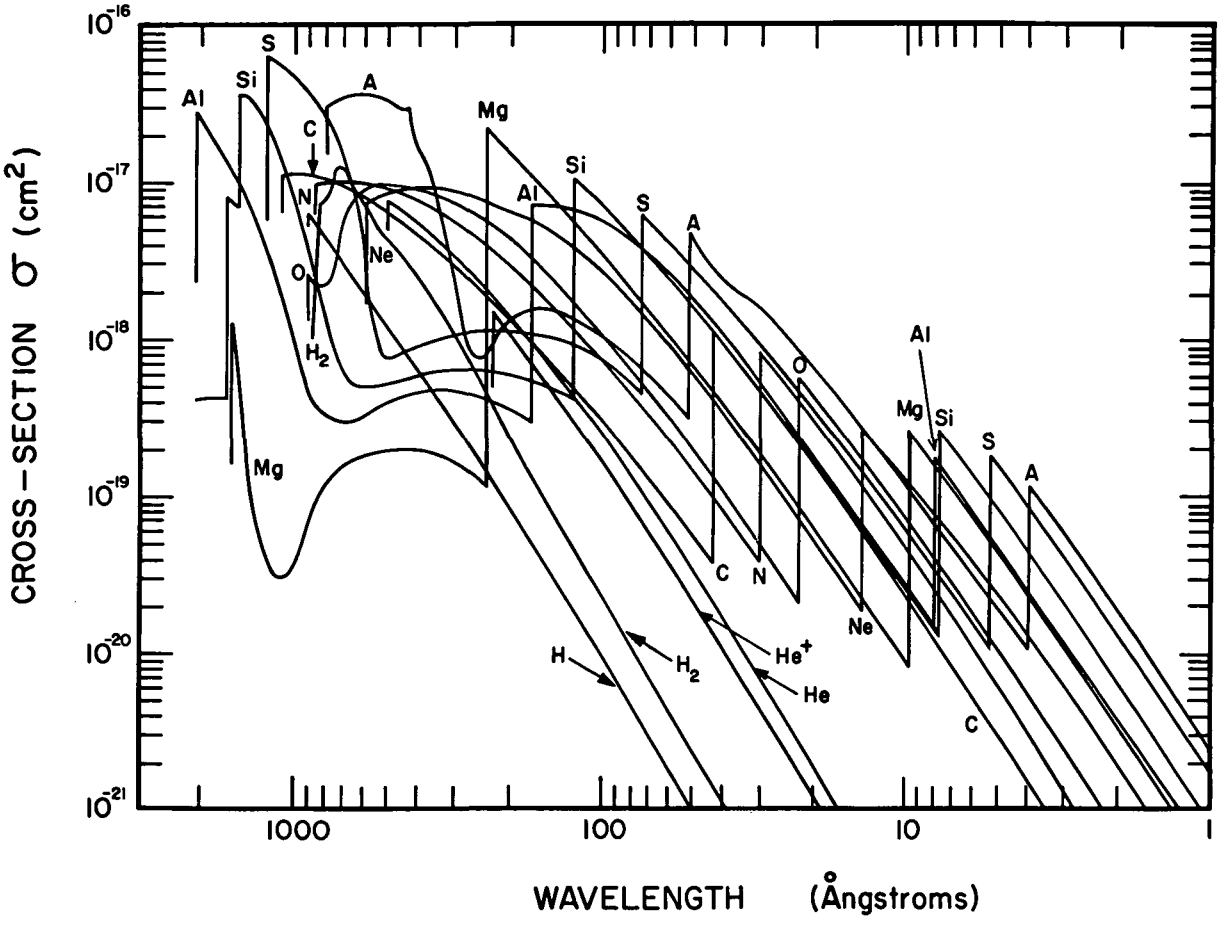



FIG. 1.—Photoabsorption cross-sections of the abundant elements in the interstellar medium as a function of wavelength

of the gas has little influence on $\sigma_e$ at soft X-ray and extreme ultraviolet wavelengths. This assumption is reasonable at X-ray wavelengths ($\sim 10$ Å), because single ionization does not change the K-shell photoabsorption cross-section of the heavy elements significantly, and these elements dominate $\sigma_e$ at wavelengths shortward of the oxygen K-edge (22.3 Å). However, at longer wavelengths most of the absorption is caused by hydrogen and helium, and as a consequence calculations of the opacity may contain errors because the degrees of ionization of these two species are unknown. In drawing figure 2, we have made the further assumption that, for the purpose of calculating $\sigma_e$ at wavelengths longward of the H I edge, the degree of ionization of the heavy elements is 99 percent.

If the interstellar medium is ionized by extreme-ultraviolet radiation (500 Å $> \lambda >$ 100 Å), the large photoionization cross-section of helium may result in it being ionized preferentially with respect to the hydrogen (Bergeron and Souffrin 1971). The cross-section of the helium ion has been plotted in figure 1, by combining the results of Allen (1963) for wavelengths between 228 and 120 Å with those of Brown and Gould (1970) for shorter wavelengths. As the cross-section of the ion is less than that of the helium atom, the effect of extreme-ultraviolet radiation might be to depress $\sigma_e$ at wavelengths below the He I edge at 504 Å. However, if the interstellar medium is ionized by far-ultraviolet radiation ($\lambda >$ 500 Å), as in an H II region, the hydrogen is ionized preferentially and the effect on $\sigma_e$ is very different.

In figure 2 we have plotted the effective cross-section in an H II region for two cases, one in which only the hydrogen is ionized and the other in which helium also is singly ionized. Only wavelengths greater than 50 Å are considered, so that the heavy elements may be ignored. Which of these conditions prevails is determined by the temperatures of the exciting stars, and Rubin (1969) has shown that the onset of helium ionization is rather sudden. The transition occurs when the class of the exciting stars is about O7. In such regions the major effect of the ionizing radiation is a

TABLE 1

SUMMARY OF THE SOURCES OF PHOTOABSORPTION CROSS-SECTIONS USED IN THIS STUDY

| Element | Wavelength (Å) | Source Reference |
|---|---|---|
| Carbon..... | 1102–44 | McGuire 1968 |
| | 23–82 | Denne 1970 |
| | 44–1 | Brown and Gould 1970 |
| Nitrogen.... | 852–600 | Comes and Elzer 1968 |
| | 600–30 | McGuire 1968 |
| | 23–82 | Denne 1970 |
| | 30–1 | Brown and Gould 1970 |
| Oxygen..... | 910–500 | Comes, Speier, and Elzer 1968 |
| | 500–124 | McGuire 1968 |
| | 124–1 | Brown and Gould 1970 |
| | 23–82 | Denne 1970 |
| Neon..... | 575–50 | Samson 1966 |
| | 23–82 | Denne 1970 |
| | 50–1 | Brown and Gould 1970 |
| Magnesium . | 1622–1500 | Peach 1970 |
| | 1500–197 | McGuire 1968 |
| | 197–10 | Henke, White, and Lundberg 1957 |
| | 10–1 | Brown and Gould 1970 |
| | | Guttman and Wagonfeld 1967 |
| Silicon..... | 1978–1500 | Peach 1970 |
| | 1500–105 | McGuire 1968 |
| | 105–8 | Henke *et al.* 1957 |
| | 8–1 | Brown and Gould 1970 |
| | | Guttman and Wagonfeld 1967 |
| Sulfur ..... | 1197–64 | McGuire 1970 |
| | 64–5 | Henke *et al.* 1957 |
| | 5–1 | Brown and Gould 1970 |
| | | Guttman and Wagonfeld 1967 |
| Aluminum .. | 2074–1500 | Peach 1970 |
| | 1500–3 | McGuire 1968 |
| Argon..... | 786–209 | Samson 1966 |
| | 250–4 | Lukirskii and Zimkina 1963 |
| | 23–82 | Denne 1970 |
| | 5–1 | Brown and Gould 1970 |
| | | Guttman and Wagonfeld 1967 |

drastic reduction of $\sigma_e$ at wavelengths longward of the He I or He II edge.

The assumed composition of the interstellar medium strongly influences the calculated value of $\sigma_e$. The presence of heavy elements, particularly oxygen, is decisive at wavelengths shortward of the oxygen K-edge (22.3 Å), but rapidly becomes negligible at longer wavelengths. Heavy elements account for 10 percent of $\sigma_e$ at 50 Å and barely 3 percent at 200 Å. Helium is responsible for about 65 percent if $\sigma_e$ is between 50 and 100 Å, beyond which its contribution decreases approximately linearly with wavelength to about 33 percent at the He I edge, 504 Å. Therefore, if the interstellar medium is inhomogeneous in composition, the value of the effective absorption cross-section will depend upon the line of sight.

Recent observations (Carruthers 1970; Spitzer *et al.* 1973) indicate that in some regions most or all of the hydrogen is in molecular form. We show in figure 1 the photoabsorption cross-section of the hydrogen molecule. This is a composite of the results cited by Cook and Metzger (1964) between 860 and 550 Å, by Samson and Cairns (1965) between 452 and 209 Å, by Denne (1970) between 82 and 23 Å, and by Craseman *et al.* (1973) at wavelengths of 2.3 and 1.5 Å. Interpolations have been made among the various experimental points. The experimental points are about 50 percent larger than twice the cross-section for atomic hydrogen. The cross-sections for molecular hydrogen shown in figure 1 differ markedly from the calculated cross-sections given by Brown and Gould (1970) between 124 and 1.24 Å (0.1 and 10 keV). Their results are greater than those shown in figure 1 by a factor 10 at 100 Å and a factor 14 at 10 Å. The curve in figure 2 shows the effective absorption cross-section of the interstellar medium when all the hydrogen is in molecular form.

TABLE 2

ASSUMED COMPOSITION OF THE INTERSTELLAR MEDIUM

| Element | Number Abundance |
|---|---|
| H ..... | 1 |
| He..... | $8.31 \times 10^{-2}$ |
| C..... | $3.98 \times 10^{-4}$ |
| N..... | $1.12 \times 10^{-4}$ |
| O..... | $8.91 \times 10^{-4}$ |
| Ne..... | $1.00 \times 10^{-4}$ |
| Mg ..... | $2.51 \times 10^{-5}$ |
| Al ..... | $1.9 \times 10^{-6}$ |
| Si..... | $3.10 \times 10^{-5}$ |
| S ..... | $2.24 \times 10^{-5}$ |
| A..... | $7.59 \times 10^{-6}$ |

The effect of interstellar dust on $\sigma_e$ has been estimated using the reddening curve between 1000 and 2000 Å obtained from measurements by OAO-2 (Bless and Savage 1972). The effective cross-section of a dust particle is related to the normalized color excess, $\epsilon(\lambda)$, by the equation

$$\sigma_d = \frac{A_V}{1.086\, n_{\rm H}} \left[ \frac{\epsilon}{R} + 1 \right],$$

where $A_V$ = absorption in magnitudes at visual wavelengths, per unit distance, $R = A_V/E_{B-V}$, and $\epsilon = E_{\lambda-V}/E_{B-V}$.

We assume $R = 3.0$, and make the simple approximation that the ratio $A_V/n_{\rm H}$ is independent of the direction of view. For example, in obscured regions we might observe $A_V \simeq 1.5\, M$ kpc$^{-1}$ and $n_{\rm H} \simeq 1$ H cm$^{-3}$, and in clear regions $A_V \simeq 0.3\, M$ kpc$^{-1}$ and $n_{\rm H} \simeq 0.2$ H cm$^{-3}$, so that in both cases the ratio is 1.5. Taking this value, we obtain

$$\sigma_d = 1.5 \times 10^{-22}(\epsilon + 3)\ \text{cm}^2 .$$

This cross-section is plotted in figure 1 as two curves, corresponding to the upper and lower bounds of the OAO-2 results for $\epsilon(\lambda)$. Little is known of the opacity of interstellar dust at wavelengths below 1000 Å. However, if the scattering is primarily Mie scattering, it

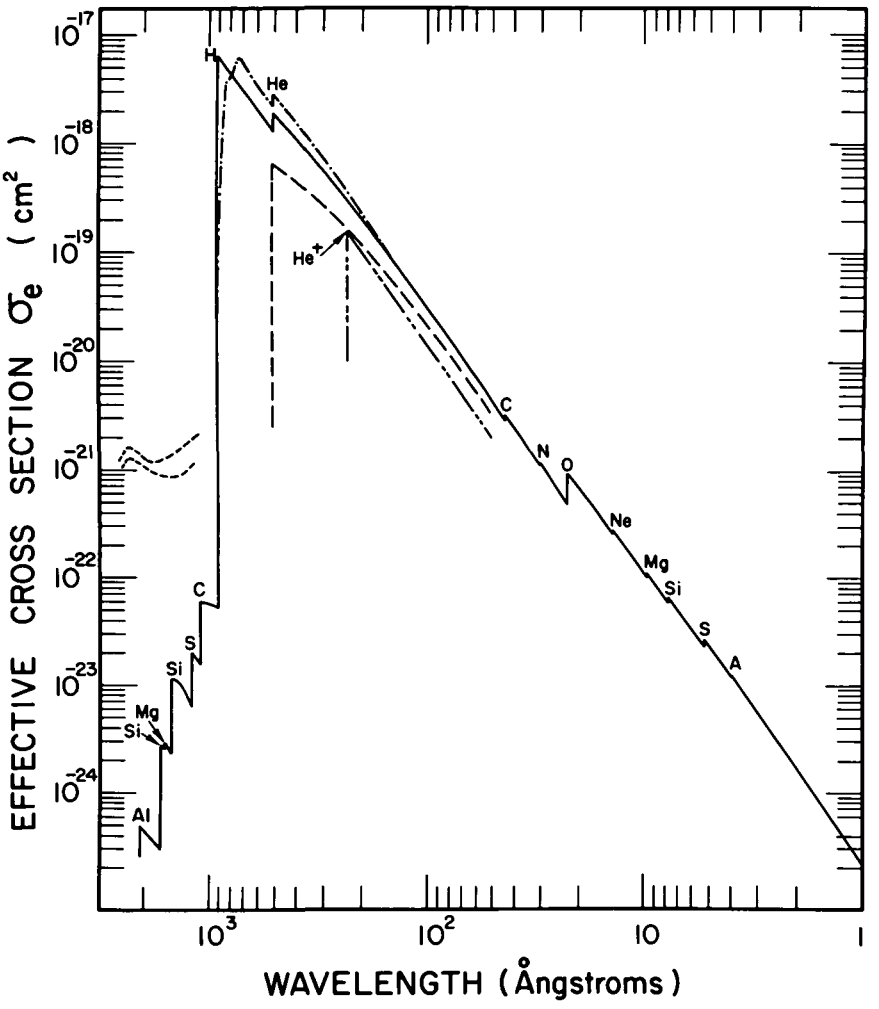



FIG. 2.—Effective cross-section (cross-section per hydrogen atom or proton) of the interstellar medium. ——— gaseous component with normal composition and temperature; —·—hydrogen, molecular form; —— —— H II region about a B star; ——-- H II region about an O star; ---- dust.

would be expected that the curves in figure 2 would flatten and become independent of wavelength at shorter wavelengths. At high photon energies, scattering by dust particles occurs mainly in the forward direction, and although the resulting effects may be observable, the opacity will not affect wide field-of-view measurements (Bowyer, Lampton, and Mack 1970). To summarize, the opacity of the interstellar medium appears to be affected significantly by dust only at wavelengths greater than 912 Å.

### III. THE OPACITY OF THE INTERSTELLAR MEDIUM

The optical depth in the interstellar medium is the product of the effective cross-section and the column density of hydrogen atoms in the line of sight. Early estimates of the optical depth relied upon measurements of 21-cm emission and absorption in the galactic plane, which were interpreted in terms of an essentially uniform gas of density $\sim 1$ cm$^{-3}$. Such a number led immediately to the forbidding view of soft X-ray and extreme-ultraviolet observations which was described by Aller (1959). This simple picture has been changed radically by the discovery that the interstellar medium is inhomogeneous and may contain extensive regions of tenuous gas.

The cloudy structure of galactic hydrogen has been known for a long time, but only recently was it realized that it might be far from isothermal. In 1968 Spitzer suggested that observations of H I regions might be interpreted in terms of cool, dense clouds in pressure equilibrium with a hot, tenuous intercloud gas, and Field, Goldsmith, and Habing (1969) proposed a two-phase medium heated by low-energy cosmic rays, in which the temperature and density were 18° K and

100 $cm^{-3}$ for the clouds and 9000° K and 0.2 $cm^{-3}$ for the intercloud gas. The clouds would contain about 75 percent of the mass but occupy less than 1 percent of the volume. Weaver (1970) and Silk (1973) have reviewed measurements of 21-cm emission and absorption in the interstellar gas, which seem to support the existence of two phases but do not verify the temperatures and densities predicted by Field *et al.* (1969).

We shall confine our attention to a region around the Sun extending to distances of about 1000 pc, which is believed (Weaver 1970) to lie in a large spur projecting from a galactic spiral arm. Heiles (1967) has examined 21-cm emission in detail from a small portion of the sky and found a few clouds with $\langle$radius$\rangle \simeq$ 19 pc, $\langle$density$\rangle \simeq 4$ $cm^{-3}$, and $\langle$column density$\rangle \simeq 2 \times 10^{20}$ $cm^{-2}$. He also found many "cloudlets" containing about 10 percent of the total cloud mass, with $\langle$radius$\rangle \simeq 2.2$ pc, $\langle$density$\rangle \simeq 2$ $cm^{-3}$, and $\langle$column density$\rangle \simeq 1.2 \times 10^{19}$ $cm^{-2}$. Radhakrishnan *et al.* (1972) have examined the 21-cm emission and absorption spectra in the direction of 35 extragalactic sources, most of which lie at intermediate and high latitudes. In these cases the gas being examined lies within 1000 pc. They conclude that about half the mass of the interstellar gas resides in clouds with column densities ranging from $2.5 \times 10^{19}$ to $10^{21}$ $cm^{-2}$. The average cloud has a column density of $3 \times 10^{20}$ $cm^{-2}$ and a temperature of 80° K, and the average spacing between clouds is about 200 pc.

Based on these two sets of observations, we adopt for further discussion an interstellar medium containing standard clouds of radius 20 pc, density 4 $cm^{-3}$, and temperature 80° K. In addition, we should consider the possibility that cloudlets of the kind found by Heiles inhabit most of the region we are considering, with a mean spacing of 20 pc. To be in pressure equilibrium with the same gas as the clouds, the cloudlets should have a mean temperature of about 170° K, a value in agreement with the velocity dispersion in cloudlets measured by Heiles. More recent observations by Heiles (1973) suggest that the dense phase of the interstellar medium is more likely to be found in filamentary structures than in clouds of a roughly spherical shape, so that the above picture may be naïve. However, this new picture of the cool hydrogen should not impede our arguments that the interstellar medium is very inhomogeneous.

The intercloud medium has been studied by Mebold (1972), who obtained 21-cm emission spectra at a galactic latitude of about 30° and at all longitudes. He concludes that the intercloud gas has a mean density of 0.2 $cm^{-3}$, a scale height of about 200 pc, and a mean velocity dispersion, after allowance has been made for line broadening due to galactic rotation, of 8.8 km $s^{-1}$. The latter figure, which agrees well with the value of 8 km $s^{-1}$ measured by Heiles (1967), places an upper limit on the gas temperature of 9400° K. A lower limit which is less than 1000° K is given by Radhakrishnan *et al.* (1972).

The contribution of turbulence to the velocity dispersion is a major stumbling block in estimating the temperature of the intercloud gas. However, Baker (1973) has analyzed statistically the 21-cm emission spectra in two regions of the sky ($l = 87.^{\circ}5$, $b = 24.^{\circ}5$, and $l = 180°$, $b = -24°$), in order to separate the turbulent and thermal velocity dispersions, and has deduced that the temperatures in these two regions are 5300° and 600° K, respectively. The latter value does not appear typical, as the velocity dispersion in the second region is lower than the mean derived from the extensive sky surveys of Mebold (1972) and Radhakrishnan *et al.* (1972).

If we assume that the density and temperature in the intercloud gas are, respectively, 0.2 $cm^{-3}$ and 5000° K, then the resulting kinetic pressure, 1000 $cm^{-3}$ ° K, is 3 times greater than the pressure in the clouds. If the clouds are not to collapse, then either the clouds must be supported internally by magnetic or turbulent pressure, and/or the values we have assumed for the density and temperature of the intercloud gas may not be typical and may be too large.

Observations of L$\alpha$ radiation lend some support to the latter conclusion. The OAO measurements (Savage and Jenkins 1972) of L$\alpha$ absorption in the spectra of O and B stars lying within 1000 pc of the Sun, suggest that in many regions of the local spur the intercloud gas density is less than 0.2 $cm^{-3}$. A histogram of values of the mean hydrogen density in the line of sight shows a maximum near 0.2 $cm^{-3}$, and the measurements may be divided approximately equally among the intervals 0–0.2, 0.2–0.4, and 0.4–1.0 $cm^{-3}$. Similar observations of 10 of the stars observed by OAO have been made by an instrument in the *Mariner 9* spacecraft. The results (Bohlin 1973) corroborate those obtained from OAO, and yield column densities which on the average are 30 percent lower. As most of the stars observed by OAO are over 200 pc from the Sun, in most cases at least one cloud should lie in the line of sight and only one-half of the column density should be attributed to the intercloud medium. Therefore, there should be extensive regions in which the intercloud density is about 0.1 $cm^{-3}$.

Information on the properties of the interstellar medium near the Sun has been provided by measurements (Thomas and Krassa 1971; Bertaux and Blamont 1971) of solar L$\alpha$ radiation scattered by interstellar atomic hydrogen streaming toward the Sun. The results have been fitted to a model (Thomas 1972), which predicts upper limits to the density and temperature of the local hydrogen. These are, respectively, 0.12 $cm^{-3}$ and 3650° K for the results of Thomas and Krassa (1971), and 0.09 and 5870° K for the results of Bertaux and Blamont (1971). The upper limit to the gas pressure is about 500 $cm^{-3}$ ° K, which is only 50 percent greater than the kinetic pressure in our standard cloud. Thomas' results imply that the solar system is moving through the intercloud gas, in a region which is less dense than the average.

We conclude from this picture of the interstellar medium that within 200 pc of the Sun there is a high probability that the mean density along the line of sight is only 0.1 $cm^{-3}$, as there are few clouds in this region and any cloudlets would cover only a few percent of the sky. At greater distances cloud will tend to make as great a contribution to the column density as the intercloud gas, and we expect to find a mean density of 0.2 $cm^{-3}$ in many directions.

In figure 3 we have used the effective cross-sections shown in figure 2 and densities of 0.1 and 0.2 $cm^{-3}$ to estimate the distance over which radiation is attenuated by 90 percent. The opacity of the ionized fraction of the intercloud gas has been ignored. We conclude that the visibility in the interstellar medium should, in many areas of the sky, be about 800 pc at 50 Å, 200 pc at 100 Å, and 10 pc at 300 Å. Observations at longer wavelengths may be restricted to objects within approximately 6 pc of the Sun, except for those view directions in which the line-of-sight column densities are unusually low.

We have found some evidence which at least encourages the hope that in a few areas of the sky unusually low column densities will be found, making possible soft X-ray and extreme ultraviolet observations over greater distances. Lα absorption measurements (Savage and Jenkins 1972) show an extended region of tenuous gas stretching along the galactic plane from Puppis to Orion. The lowest mean density measured in this region was 0.03 (+0.03, −0.015) $cm^{-3}$, along the line of sight to β CMa (a distance of 220 pc). In addition, the absence of Lα absorption in the spectrum of Arcturus (Rottman *et al.* 1971), which is 11 pc from the Sun, implies an upper limit to the mean hydrogen density in this direction of 0.05 $cm^{-3}$. More recent results, obtained by OAO-3, suggest average densities along the lines of sight to α Leo, α Eri, and ν Sco of 0.02, 0.07, and 0.05 $cm^{-3}$, respectively (Rogerson *et al.* 1973).

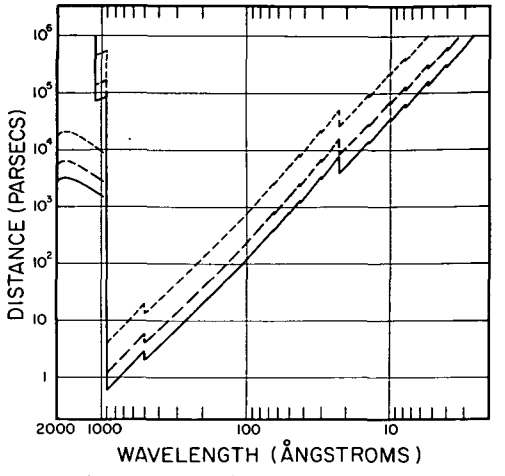



FIG. 3.—Distance at which attenuation of the incident radiation reaches 90 percent, as a function of wavelength. An un-ionized interstellar medium of normal composition is assumed, and the contribution of the dust is based on the upper curve in fig. 2. —— $\eta_H = 0.2\,cm^{-3}$; — — $\eta_H = 0.1\,cm^{-3}$; – – – $\eta_H = 0.03\,cm^{-3}$.

A number of observations suggest that the interstellar medium may be disturbed violently, and consequently the intercloud gas may be quite inhomogeneous. Baker (1973) found very different temperatures in the two regions of the sky which he analyzed, and this implies that the ratio of their mean densities may be as great as a factor of 5. Further, in both regions he found a turbulent velocity dispersion so large that the randomly moving clumps of gas must be moving supersonically, which suggests that the intercloud gas is being stirred continuously and dissipating energy through shock waves. These are not peculiar results, for large velocity dispersions in the intercloud gas are a persistent feature of 21-cm emission measurements (Heiles 1967; Mebold 1972; Radhakrishnan *et al.* 1972). The energy-dissipation rate is large, about $3 \times 10^{-26}$ ergs $cm^{-3}$ $s^{-1}$, and falls within the range of values which have been invoked in explanations of the heating and ionization of the interstellar medium. The range extends from $\sim 3 \times 10^{-27}$ ergs $cm^{-3}$ $s^{-1}$ (Field *et al.* 1969) to $10^{-25}$ ergs $cm^{-3}$ $s^{-1}$ (Silk 1973). Baker finds evidence that the larger elements in the turbulence participate in a streaming motion directed toward the Sun, and suggests that such streams may be the source of the energy. Such massive streams, moving in the galactic plane at velocities which approach 100 km $s^{-1}$ in some instances and usually are negative, have been noted by Weaver (1970). A possible cause of these streams is convective heat transfer between the interstellar medium and a surrounding layer of hot, tenuous gas.

In a recent paper Silk (1973) has argued that evidence for a low degree of ionization in clouds makes it difficult to accept low-energy cosmic rays as the primary source of ionization in the interstellar medium. If this were true, then heating and ionization in the interstellar medium would be confined essentially to H II regions in the intercloud gas. Alternative sources of heating and ionization in the intercloud gas are extreme-ultraviolet radiation emitted by supernova blast waves or ultraviolet stars, or convective heat transfer with a hot, tenuous halo surrounding the cooler hydrogen in the galactic disk. Such processes should result in irregular distributions of density and temperature in the intercloud gas, and in certain regions ionization of hydrogen and helium might reduce significantly the opacity of the gas at soft X-ray and extreme ultraviolet wavelengths.

Silk (1973) argues further that pulsar dispersion, and free-free emission and absorption of the radio continuum, might be ascribed for the most part to the electrons of the intercloud gas. This would call for a high ionization rate ($\zeta \simeq 10^{-14}$ $s^{-1}$), and for the gas to be as much as 50 percent ionized. Care must be taken in comparing radio-continuum with pulsar-dispersion

estimates of electron density, for the former gives an estimate of $\langle N_e^2 \rangle$ and the latter $\langle N_e \rangle$, and it is possible that free-free absorption and emission are dominated by H II regions around hot stars (Gould 1971). However, pulsar dispersion appears to be determined by the intercloud gas, and after analyzing the dispersions of 14 pulsars with latitudes greater than 25°, we find that within a few hundred parsecs of the Sun the electron density lies between 0.05 and 0.1 $cm^{-3}$. When these values are compared with the values we have ascribed to the atomic hydrogen density in the local intercloud gas, 0.1–0.2 $cm^{-3}$, a high degree of ionization is suggested.

While it is difficult to obtain a distinct picture of the intercloud gas from 21 cm, radio-continuum, pulsar-dispersion, solar Lα scattering, and stellar Lα absorption measurements, it appears equally difficult to see it as a placid, uniform continuum surrounding H I regions. There is strong evidence that it is heated, ionized, and disturbed violently by agents which produce significant inhomogeneities in density and temperature.

In particular, we draw attention to the sector of the Galaxy between longitudes of 180° and 270°, in which Lα absorption measurements over path lengths as long as 900 pc have yielded extraordinarily low values for the mean hydrogen density. Seven of the 14 high-latitude pulsars ($|b| > 25°$) which have been discovered are clustered in this sector (we include MP 0256 at $l = 271°$), from which we deduce an electron density in the plane of 0.05 $cm^{-3}$. This is confirmed by the dispersion of the Crab pulsar, whose distance is known well. A value more than twice as large is found for the remaining seven pulsars, which are at longitudes between 0 and 140°. This part of the Galaxy may turn out to be just an interarm region between the Sagittarius arm and the local spur, but other observations, such as the soft X-ray excess at low latitudes reported by Bunner *et al.* (1973), Yentis, Novick, and Vanden Bout (1972), and Garmire (1973), and the large jet of high-velocity hydrogen found at $l = 196°.5$ by Weaver (1970), may be evidence of something more unusual.

Therefore, as there may well exist regions in the sky where the mean density along the line of sight is appreciably less than 0.1 $cm^{-3}$, we show in figure 3 the 90 percent attenuation distance in a gas of density 0.03 $cm^{-3}$, which is the lowest value obtained from the OAO observations (Savage and Jenkins 1972). Under such conditions the visibility at 100 Å would improve to nearly 800 pc, and at 300 Å to 40 pc.

Finally, observations of the X-ray background flux may prove very instructive. If this radiation is produced outside the galactic hydrogen layer, observations at wavelengths between 50 and 100 Å may reveal clouds or other structures in the interstellar medium. If it is produced within the layer, observations between 100 and 300 Å should reveal its origin. In either case, the position of a peak or plateau in the spectrum is of particular interest, for it may provide some insight into the sources of the soft X-ray background.

## REFERENCES


Allen, C. W., 1963, *Astrophysical Quantities* (London: Athlone Press).
Aller, L. H., 1959, *Pub. A.S.P.*, **71**, 324.
Baker, P. L., 1973, *Astr. and Ap.*, **23**, 81.
Bell, K. L., and Kingston, A. E., 1967*a*, *M.N.R.A.S.*, **136**, 24.
———. 1967*b*, *Proc. Phys. Soc.*, **90**, 31.
Bergeron, J., and Souffrin, S., 1971, *Astr. and Ap.*, **11**, 40.
Bertaux, J. L., and Blamont, J. E., 1971, *Astr. and Ap.*, **11**, 200.
Beynon, J. D. E., and Cairns, R. B., 1965, *Proc. Phys. Soc.*, **86**, 1343.
Bless, R. C., and Savage, B. D., 1972, *Ap. J.*, **171**, 293.
Bohlin, R. C., 1973, *Ap. J.*, **192**, 139.
Bowyer, C. S., Lampton, M. L., and Mack, J. E., 1970, *Nature*, **225**, 1125.
Brown, R. L., and Gould, R. J., 1970, *Phys. Rev. D*, **1**, 2252.
Bunner, A. N., Coleman, P. L., Kraushaar, W. L., McCammon, D., and Williamson, F. O., 1973, *Ap. J.*, **179**, 781.
Carruthers, G. R., 1970, *Ap. J. (Letters)*, **161**, L81.
Comes, F. J., and Elzer, A., 1968, *Zs. f. Naturforschung*, **23a**, 133.
Comes, F. J., Speier, F., and Elzer, A., 1968, *Zs. f. Naturschung*, **23a**, 125.
Cook, G. R., and Metzger, P. H., 1964, *J. Opt. Soc. Am.*, **54**, 968.
Crasemann, B., Koblas, P. E., Wang, T., Birdseye, H. E., and Chen, M. S., 1973, in preparation for *Phys. Rev. A*.
Denne, D. R., 1970, *J. of Phys. D*, **3**, 1392.
Felten, J. E., and Gould, R. J., 1966, *Phys. Rev. Letters*, **17**, 401.
Field, G. B., Goldsmith, D. W., and Habing, H. J., 1969, *Ap. J. (Letters)*, **155**, L149.
Garmire, G., 1973, private communication.
Gould, R. J., 1971, *Ap. and Space Sci.*, **10**, 265.
Guttman, A. J., and Wagonfeld, H., 1967, *Acta Crystal.*, **22**, 335.
Heiles, C., 1973, private communication.
Heiles, C., 1967, *Ap. J. Suppl.*, **15**, 97.
Henke, B. L., White, R., and Lundberg, B., 1957, *J. Appl. Phys.*, **28**, 98.
Karzas, W. J., and Latter, R., 1961, *Ap. J. Suppl.*, **6**, 167.
Lowry, J. F., Tomboulian, D. H., and Ederer, D. L., 1965, *Phys. Rev.*, **137**, A1054.
Lukirskii, A. P., and Zimkina, T. M., 1963, *Isv. Akad. Nauk USSR, Ser. Fiz.*, **27**, 817.
Lukirskii, A. P., Brytov, I. A., and Zimkina, T. M., 1964, *Opt. Spectrosc.*, **17**, 234.
McGuire, E. J., 1968, *Phys. Rev.*, **175**, 20.
Mebold, U., 1972, *Astr. and Ap.*, **19**, 13.
Palmer, P., Zuckerman, B., Penfield, M., Lilley, A. E., and Metzger, P. G., 1969, *Ap. J.*, **156**, 887.
Peach, G., 1970, *Mem. R.A.S.*, **73**, 1.
Radhakrishnan, V., Murray, J. D., Lockhart, P., and Whittle, R. P. J., 1972, *Ap. J. Suppl.*, **203**, 15.
Rogerson, J. B., York, D. G., Drake, J. F., Jenkins, E. B., Morton, D. C., and Spitzer, L., 1973, *Ap. J. (Letters)*, **181**, L110.
Rottman, G. J., Moos, H. W., Berry, J. R., and Henry, R., 1971, Johns Hopkins University Report.
Rubin, R. H., 1969, *A.J.*, **74**, 994.
Samson, J. A. R., 1966, *Advances in Atomic and Molecular Physics*, Vol. 2, ed. D. R. Bates and I. Estermann (New York: Academic Press), p. 117.
Samson, J. A. R., and Cairns, R. B., 1965, *J. Opt. Soc. Am.*, **55**, 1935.

Savage, B. D., and Jenkins, E. B., 1972, *Ap. J.*, **172**, 491.
Silk, J., 1973, submitted to *Ap. J.*
Spitzer, L., 1968, *Nebulae and Interstellar Matter*, ed. B. M. Middlehurst and L. H. Aller (Chicago: University of Chicago Press), p. 1.
Spitzer, L., Drake, J. F., Jenkins, E. B., Morton, D. C., Rogerson, J. B., and York, D. G., 1973, *Ap. J. (Letters)*, **181**, L116.
Thomas, G. E., 1972, NASA SP-308, 668.
Thomas, G. E., and Krassa, R. F., 1971, *Astr. and Ap.*, **11**, 218.
Weaver, H. F., 1970, *Interstellar Gas Dynamics, Proc. 39th IAU Symp.* ed. H. Habing, p. 22.
Yentis, D. J., Novick, R., and Vanden Bout, P., 1972, *Ap. J.*, **177**, 375.

A74-23319

(Reprinted from Nature, Vol. 247, No. 5442, pp. 513–516, February 22, 1974)

# Cosmic Far Ultraviolet Background

**Arthur Davidsen, Stuart Bowyer and Michael Lampton**

Space Sciences Laboratory and Astronomy Department, University of California, Berkeley, California 94720

*Observations of a diffuse ultraviolet background at high galactic latitude might reveal the existence of an intergalactic medium. Although existing results provide only weak constraints, current techniques are sufficient to set interesting limits on its density and temperature.*

INTEREST in the diffuse far ultraviolet background stems mainly from its relevance to the widely held view that intergalactic space may be filled with a dilute gas containing the majority of the mass of the universe. Observations which might reveal the existence of the intergalactic medium (IGM) have been attempted at radio, optical, far ultraviolet and X-ray wavelengths (see ref. 1 for a recent review). If the IGM has a density comparable to the critical density required to close the universe, then 21 cm radio observations and optical observations of QSO spectra imply the gas is highly ionised and therefore $T \gtrsim 10^4$ K. On the other hand observations of the diffuse X-ray background show that the temperature of such a gas is probably less than about $3\times10^8$ K. In the lower part of this acceptable range of temperatures, the IGM would radiate primarily in the far ultraviolet. Alternatively, the high degree of ionisation of the gas may be maintained by a large ultraviolet flux from some as yet unidentified source. Thus, the far ultraviolet region of the spectrum deserves considerable attention in the effort to determine the density and temperature of the IGM.

Since the first report[2] of an observation of the diffuse far ultraviolet background, there have been several measurements attempted, both longward and shortward of Lyman $\alpha$. As the reported intensities range over nearly three orders of magnitude, a critical evaluation is required before the observations may be interpreted in terms of limits on the physical state of the IGM. Here we discuss the expected intensities of various possible components of the far ultraviolet background and review the relevant observations.

We conclude that existing results do not place interesting constraints on the density of the IGM. Current techniques and instrumentation for far ultraviolet astronomy are, however, sufficient to achieve vastly improved limits, but new observations are required to determine whether the IGM can be detected in the far ultraviolet or whether the extragalactic component of the background is masked by radiation with a more local origin.

There are several contributors to the far ultraviolet background radiation in addition to the possibility of emission from the IGM. The extragalactic background will include a contribution from galaxies and QSOs. A galactic background will be present which will consist of direct stellar radiation and a truly diffuse component resulting from starlight scattered by interstellar dust. Locally there will be contributions from the zodiacal light as well as airglow emissions from the Earth's geocorona. In what follows, we discuss briefly what is known, either theoretically or observationally, about each of the components of the far ultraviolet background.

## Intergalactic medium

Attempts[3] to observe Lyman $\alpha$ absorption in the optical spectra of large redshift QSOs have led to the conclusion that the IGM is either remarkably empty, or it was highly ionised at $z \geq 2$. Such a high degree of ionisation implies temperatures greater than about $10^4$ K. Kurt and Sunyaev[2] first discussed Lyman $\alpha$ radiation from recombinations in the IGM. Emission in this line is maximised at $T \approx 2\times10^4$ K, where an intensity of $4\times10^{-10}$ erg (cm$^2$ s sr Å)$^{-1}$ just longward of Lyman $\alpha$ is produced in a uniform universe having the critical density and with the Hubble constant $H_0=50$ km s$^{-1}$ Mpc$^{-1}$. The best current estimate[4] of the deceleration parameter in a Friedmann model of the universe is $q_0=1\pm1$ and this yields an intensity of $1.6\times10^{-9}$ erg (cm$^2$ s sr Å)$^{-1}$.

Silk and Tarter[5] have considered a clumpy IGM, consisting of clouds of ionised gas which are distributed similarly to galaxies. In such a model, it is possible to gravitationally bind groups and small clusters of galaxies, as well as the universe, without violating limits on the ultraviolet and soft X-ray backgrounds. Since the clumpiness increases the emissivity of the IGM per unit mass, a substantial flux of Lyman $\alpha$ recombination radiation may be produced even at $T \gg 2\times10^4$ K.

The cosmic Lyman $\alpha$ flux is also relevant to the arguments which lead to the conclusion that a dense, neutral IGM cannot exist. Although the QSO observations already mentioned provide extremely strong limits on neutral hydrogen ($n_H$ $(z=2)<1.5\times10^{-11}$ cm$^{-3}$), this result depends on the assumption that QSO redshifts are cosmological. Even with this assumption, these data provide information only about a distant epoch. Since the cosmological distance assumption is questioned by some astronomers and since the ionisation of the IGM at the present epoch may differ from that at $z\simeq2$, the limits on neutral hydrogen obtained through 21 cm observations of low redshift objects are still of interest. These observations[1] give $n_H/T_E<1.8\times10^{-8}$ cm$^{-3}$ K$^{-1}$, where $T_E$ is the excitation temperature of the 21 cm line. The principal unknown contribution to $T_E$ comes from recombination Lyman $\alpha$ in the IGM. Thus an upper limit on the Lyman $\alpha$ flux is required to set a limit on $T_E$ and thus on $n_H$. An indirect limit of $T_E<12$ K has been set by Field[1]. A direct observational confirmation of this requires the establishment of an upper limit of $4\times10^{-10}$ erg (cm$^2$ s sr Å)$^{-1}$ for the cosmic Lyman $\alpha$ flux.

In addition to ionised hydrogen, the IGM is generally believed to contain about 10% helium which was produced in the primeval fireball. At a temperature of about $10^5$ K the IGM will thus radiate strongly in the HeII $\lambda$ 304 Å recombination line. Radiation of this wavelength, originating at epochs with $z\simeq2.5$ to 3.0, will be redshifted into the observable spectral region around 1,100 Å. Kurt and Sunyaev[2,6] first pointed out that because the density in an expanding universe varies as $(1+z)^3$ and the intensity of a recombination line varies as the square of the density, measurement of the redshifted helium radiation provides an even more sensitive probe of the IGM than that provided by the hydrogen radiation. Thus observations shortward as well as longward of Lyman $\alpha$ may yield information on the thermal history of the IGM back to epochs where the most distant QSOs have been observed.

## Galaxies and QSOs

The possibility that the IGM is photoionised by far UV radiation from QSOs and galaxies has been considered by several authors[7–11]. Rees and Setti[9] estimate that QSOs alone

would produce a Lyman limit intensity of about $4\times10^{-11}$ to $15\times10^{-11}$ erg (cm$^2$ s sr Å)$^{-1}$. They also suggest a similar contribution could be made by an extrapolation of the diffuse X-ray background spectrum to ultraviolet wavelengths. Such a flux would be sufficient to maintain a high degree of ionisation, consistent with the Gunn–Peterson result, in an IGM whose density is $n\sim10^{-7}$ cm$^{-3}$. Rees and Setti conclude that an IGM with the critical density could be sufficiently photoionised by a far ultraviolet background whose intensity is $\sim2\times10^{-9}$ erg (cm$^2$ s sr Å)$^{-1}$. The existence of such a universal ultraviolet flux might have profound implications beyond its effect on the IGM.

A suggestion that normal galaxies are strong emitters of far ultraviolet radiation has recently been made by Hills[12]. According to this work, ultraviolet dwarfs with temperatures of about $2\times10^5$ K emit enough ultraviolet photons to maintain a high degree of ionisation in an IGM having the critical density. This radiation, however, is strongly concentrated beyond the Lyman limit, and the intensity in the observable region at approximately 1,300 Å would be only about $\sim5\times10^{-12}$ erg (cm$^2$ s sr Å)$^{-1}$. On the other hand, if such hot stars were numerous in young galaxies at large $z$, their contribution to the intensity at approximately 1,300 Å might be substantially larger.

Although the intensity of the cosmic far ultraviolet background is not predictable, it is clear that a sensitivity of $\leq10^{-9}$ erg (cm$^2$ s sr Å)$^{-1}$ should allow some interesting conclusions to be drawn. An extragalactic background of this intensity, however, might be overwhelmed by galactic and even more local sources of ultraviolet radiation, thereby forever precluding investigation of the IGM at ultraviolet wavelengths. We turn now to a consideration of these other components of the ultraviolet background.

## Galactic starlight

It is clear that observations made with detectors with large fields of view will include a spurious flux component arising from point sources whose space density is sufficiently large. The mean intensity of direct starlight in the solar neighbourhood at approximately 1,400 Å is about $1\times10^{-7}$ erg (cm$^2$ s sr Å)$^{-1}$ according to calculations by Habing[13]. Other calculations[14] suggest a variation of a factor of about ten in intensity of direct starlight from galactic plane to pole. Thus, observations made with large field detectors may be expected to yield I(1,400 Å)$\geq10^{-7}$ erg (cm$^2$ s sr Å)$^{-1}$ to within a factor of ten. This is consistent with the observations discussed below.

To observe the cosmic background, this pseudo background must be substantially reduced by using a small field of view. As the field of view is made smaller, sources of progressively higher space density are eliminated. This has a particularly important effect on far ultraviolet observations, as the relatively low density objects which may be completely eliminated with a small field of view are just those hot stars which contribute most to the ultraviolet flux measured with a large field detector. Of course, as the field of view is reduced, so is the sensitivity of the detector to diffuse radiation. There must therefore be an optimum field of view which is just small enough to eliminate a substantial contribution from stars.

For a field of view of 0.1 square degrees, only stars of spectral type F and later are numerous enough to contribute to the diffuse flux at high galactic latitude, but such stars are intrinsically weak in the far ultraviolet. If a telescope is oriented to avoid stars brighter than $V=10$, we expect the stellar contribution to the background will be less than approximately $10^{-11}$ erg (cm$^2$ s sr Å)$^{-1}$, substantially below the predicted extragalactic flux. Similar conclusions have been reached by Kurt and Sunyaev[6] and by Smirnov[15]. With this field of view the ultraviolet dwarfs discussed above will not be sufficiently numerous to produce a significant diffuse component either.

## Diffuse galactic light

Extinction of far ultraviolet radiation in the galaxy is due to interstellar grains. Some models of these grains, such as dielectric particles and Platt particles, predict high albedos. If the extinction is mainly due to scattering, a diffuse galactic light will exist, whose origin is in starlight. If the albedo is very high, the energy density of this diffuse light may be comparable to that of the direct starlight. The intensity of diffuse light will vary with direction in the Galaxy in a way which depends on the distribution of hot stars, the distribution of the dust, the albedo, and the phase function for the scattering. For a simple model of the galaxy in which the stars and dust are similarly distributed in plane parallel layers, the variation of intensity of diffuse light with galactic latitude has been calculated by Van de Hulst and de Jong[14]. Taking the grain parameters predicted by Gilra's models[16] and a galactic half optical thickness of 0.2, we find from Van de Hulst and de Jong's work that the diffuse light at the pole is about 0.5% of the total intensity (including starlight) near the galactic plane. This intensity[17] is about $5\times10^{-7}$ erg (cm$^2$ s sr Å)$^{-1}$ at 1,400 Å. Thus, the diffuse galactic light predicted at the pole in this model is $2.5\times10^{-9}$ erg (cm$^2$ s sr Å)$^{-1}$, which is somewhat larger than the expected Lyman $\alpha$ flux from the IGM.

If the plane-parallel model were applicable, the diffuse galactic light would make a measurement of the cosmic far ultraviolet background difficult. It has been shown, however, that this model for the distribution of interstellar dust is inadequate. A wide variety of studies[18] suggest that the interstellar reddening $E(B-V)=0.00$ for galactic latitudes greater than $b=50°$. This result may easily be reconciled with the usual cosecant $b$ dependence of the plane-parallel galaxy models when it is realised that the dust is concentrated in clouds. At low latitudes, where a line of sight intersects many clouds, the uniform distribution assumption is not inappropriate but at high latitudes most lines of sight simply do not intersect a dust cloud. Thus, for most directions above $b=50°$ no diffuse galactic light is expected and therefore the cosmic UV background will not be masked.

## Zodiacal light

The zodiacal light results from solar radiation scattered by interplanetary dust and is one of the principal contributors to the sky brightness at visual wavelengths. Because of the sharp decrease of the solar flux toward short wavelengths, however, it may be expected that the zodiacal light will not contribute much to the far ultraviolet background. Thus, Kurt and Sunyaev[6] concluded that the intensity of the zodiacal light at 1,300 Å is less than $2\times12^{-12}$ erg (cm$^2$ s sr Å)$^{-1}$. On the other hand, Lillie[19] presented preliminary evidence that the zodiacal light spectrum turns up below 2,500 Å. His data suggested a continued steep rise toward the shortest wavelengths observed (about 1,680 Å). More recent data (C. F. Lillie, unpublished), obtained with the ultraviolet spectrometer aboard Apollo 17, show that the spectrum turns down again below 2,000 Å. In fact, the relative intensity of the zodiacal light, normalised to the Sun, is now thought to be the same at about 1,680 Å and 4,250 Å. Thus, the conclusion of Kurt and Sunyaev that this component is negligible in the far ultraviolet is probably correct. In any case, a zodiacal light flux would certainly show a concentration toward the ecliptic. The zodiacal light, therefore, probably does not present any important limitation on the measurement of a cosmic far ultraviolet background.

## Airglow

Probably the most important limitation on rocket observations of the diffuse far ultraviolet background is the contribution of line emission originating in the geocorona. By far the strongest of these lines is Lyman $\alpha$ λ1,216, whose intensity in the night sky is $\sim10^9$ photons (cm$^2$ s sr)$^{-1}$. No other

lines have been observed, but it is possible that the OI λ1,304 triplet and OI λ1,356 will be significant at the flux levels under discussion here. If these lines have intensities of as little as 1 rayleigh, their contribution to the diffuse background measured with a typical photometer having a bandpass of about 200 Å is approximately $10^{-8}$ erg (cm² s sr Å)$^{-1}$. At present, such an intensity cannot be ruled out. Just as a small field of view is required to eliminate discrete spatial points (stars) which produce continuous spectral contributions, so is a narrow bandpass required to eliminate discrete spectral points (airglow lines) which produce a continuous spatial contribution, and attempts to observe the cosmic far ultraviolet background from rockets will require at least moderate resolution. Observations from high altitude satellites or interplanetary probes will not be limited in this way.

## Observations

In spite of several attempts to measure the intensity of the diffuse ultraviolet background just longward and shortward of Lyman α, the subject remains in a state of confusion. Results of all previous measurements, displayed in Figure 1, range over two and a half orders of magnitude. The large range of the positive detections probably results from the large fields of view employed and the distribution of hot stars in the Galaxy.

The first attempt to measure the far ultraviolet background in the neighbourhood of Lyman α was made by Kurt and Sunyaev[2] using a detector aboard the interplanetary probe Venera-3. With a field of view of approximately 300 square degrees and a bandpass from λ1,225 to 1,340 Å, they detected no signal above that due to the cosmic ray background. Kurt and Sunyaev thus obtained an upper limit $I_\nu \lesssim 5 \times 10^{-22}$ erg (cm² s sr Hz)$^{-1}$ or $I_\lambda \lesssim 1.0 \times 10^{-9}$ erg (cm² s sr Å)$^{-1}$. This result is generally considered suspect, because Kurt and Sunyaev[20] later reported a detection, with the same instrument, of the hot stars near the galactic plane at an intensity which is two orders of magnitude below the theoretically expected value[13,21] and the value observed by Hayakawa, Yamashita, and Yoshioka[17] in the same direction.

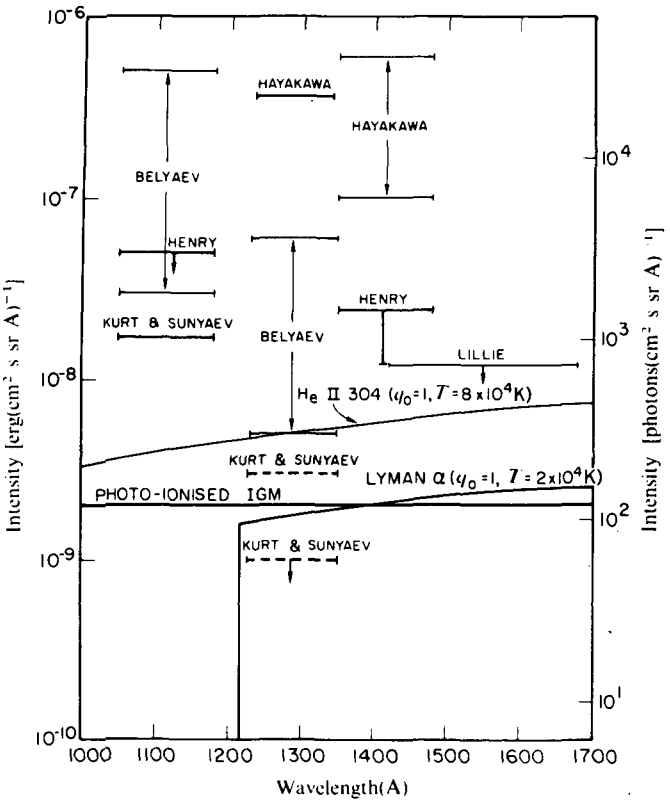



**Fig. 1** Observed and predicted intensities of the diffuse far ultraviolet background. The measurements are discussed in the text. The lowest intensities, due to Kurt and Sunyaev, are contradicted by later results. The Lyman α flux from recombinations IGM with $q_0 = 1$ and $T = 2 \times 10^4$ K (assumed constant with $Z$) is shown. A similar flux due to He II 304 Å recombinations at large $Z$ is also shown. Also presented is an estimate of the flux which is necessary to produce a sufficient degree of photoionisation in an IGM having the critical density. Local and Galactic components of diffuse light are probably weak enough at high latitude to allow detection of a cosmic flux of approx $10^{-9}$ erg (cm² s sr Å)$^{-1}$. A sensitivity at this level can be achieved with current instrumentation.

Another measurement in the region λλ1,225 to 1,340 Å was reported by Belyaev *et al.*[22] with an instrument aboard Venera-5 and Venera-6. These authors also measured the background in the range λλ1,050 to 1,180 Å. The fields of view were approximately 120 and 0.4 square degrees for the shorter and longer wavelengths, respectively. Intensities obtained in the λλ1,225 to 1,340 Å interval varied from about $2 \times 10^{-8}$ to about $5 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$ near the galactic plane and an upper limit of $I_\lambda \lesssim 5 \times 10^{-9}$ erg (cm² s sr Å)$^{-1}$ was obtained in an undetermined direction. The results were of very low statistical accuracy because of the small field of view, a rather small collecting area (about 10 cm²), and a high cosmic-ray background. In the interval λλ1,050 to 1,180 Å, Belyaev *et al.* obtained intensities ranging over about $3 \times 10^{-8}$ to $65 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$. The authors attribute the much higher intensities observed in this band to the larger field of view.

Hayakawa, Yamashita and Yoshioka[17] have measured the background intensity in the range λλ1,350 to 1,480 Å with a detector whose field of view was 20 square degrees. Their results ranged from about $5 \times 10^{-7}$ erg (cm² s sr Å)$^{-1}$ near the galactic plane smoothly down to about $1 \times 10^{-7}$ erg (cm² s sr Å)$^{-1}$ at $b \approx 40°$. These authors also obtained an upper limit to a possible extragalactic intensity of $\leq 1.7 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$ on the basis of an analysis of the latitude variation in terms of the scattering properties of interstellar grains. This analysis was subsequently revised[23] with the result that no information on the extragalactic component was obtained.

Using photometers aboard OAO-2, Lillie[24] has obtained an upper limit on the diffuse intensity at high galactic latitude in a 260 Å band centred at 1,550 Å. With the extremely small field of view (about 0.02 square degrees) of this instrument, the contribution due to hot stars was eliminated. Unfortunately Lillie observed no signal above the dark counting rate because of the low throughput of the instrument. His resultant upper limit was $1.2 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$.

Henry[25] has reported measurements of the high latitude background in the range λλ1,050 to 1,180 Å and λλ1,350 to 1,475 Å, made with a field of view of approximately 100 square degrees. In the longer wavelength range, he obtained an intensity of about $2 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$. Shortward of Lyman α, he obtained only the upper limit ($I_\lambda$ 1,100 Å) $\lesssim 3 \times 10^{-8}$ erg (cm² s sr Å)$^{-1}$ since the measured signal was largely due to residual sensitivity of his detector to Lyman α radiation.

## Intensities

It seems likely that most of the large range of intensities in Fig. 1 is due to the varying number of hot stars with galactic latitude. These are the main contributors to the intensity measured with a large field of view. In addition there is undoubtedly an important contribution from diffuse galactic light at low latitudes. Except for the results of Kurt and Sunyaev, the lowest intensities were obtained with the detectors having the smallest fields of view. These results were upper limits, and are limited by the sensitivity of the instruments. Unfortunately, these limits are not yet strong

enough to place important constraints on the density of the IGM. A sensitivity of $\lesssim 10^{-9}$ erg (cm$^2$ s sr Å)$^{-1}$, achievable with current instrumentation, would provide an interesting limit. Local and galactic components of diffuse far ultraviolet radiation are probably weak enough at high galactic latitude to allow detection of a cosmic intensity of this magnitude.

*This work has been supported by a* NASA *grant.*


[1] Field, G. B., *A. Revs. Astr. Astrophys.*, **10**, 227 (1972).
[2] Kurt, V. G., and Sunyaev, R. A., *JETP Lett.*, **5**, 246 (1967).
[3] Gunn, J., and Peterson, B., *Astrophys. J.*, **142**, 1633 (1965).
[4] Sandage, A. R., *Astrophys. J.*, **178**, 1 (1972).
[5] Silk, J., and Tarter, J., *Astrophys. J.*, **183**, 387 (1973).
[6] Kurt, V. G., and Sunyaev, R. A., *Cosmic Research*, **5**, 496 (1967).
[7] Weymann, R., *Astrophys. J.*, **147**, 887 (1967).
[8] Arons, J., and McCray, R., *Astrophys. Lett.*, **5**, 123 (1969).
[9] Rees, M. J., and Setti, G., *Astr. Astrophys.*, **8**, 410 (1970).
[10] Bergeron, J., and Salpeter, E., *Astrophys. Lett.*, **7**, 115 (1970).
[11] Arons, J., and Wingert, D. W., *Astrophys. J.*, **177**, 1 (1972).
[12] Hills, J. G., *Astrophys. Lett.*, **12**, 9 (1972).
[13] Habing, H. J., *Bull. Astr. Inst. Neth.*, **19**, 421 (1968).
[14] Van de Hulst, H. C., and de Jong, T., *Physica*, **41**, 151 (1969).
[15] Smirnov, A. S., *Cosmic Research*, **8**, 685 (1971).
[16] Gilra, D. P., *Nature*, **229**, 237 (1971).
[17] Hayakawa, S., Yamashita, K., and Yoshioka, S., *Astrophys. Space Sci.*, **5**, 493 (1969).
[18] Sandage, A. R., *Astrophys. J.*, **183**, 711 (1973).
[19] Lillie, C. F., *Scientific Results from OAO-2*, 95 (NASA, 1972).
[20] Kurt, V. G., and Sunyaev, R. A., *Soviet Phys. Astr.*, **11**, 928 (1968).
[21] Witt, A. N., and Johnson, M. W., *Astrophys. J.*, **181**, 363 (1973).
[22] Belyaev, V. P., Kurt, V. G., Melioranskii, A. S., Smirnov, A. S., Sorokin, L. S., and Tiit, V. M., *Cosmic Research*, **8**, 677 (1971).
[23] Yoshioka, S., *Astrophys. Space Sci.*, **16**, 421 (1972).
[24] Lillie, C. F., *Scientific Results from OAO-2*, 583 (NASA Publication, 1972).
[25] Henry, R. C., *Astrophys. J.*, **179**, 97 (1973).


Printed in Great Britain by Henry Ling Ltd., at the Dorset Press, Dorchester, Dorset

THE ASTROPHYSICAL JOURNAL, 188:L71–L73, 1974 March 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

A74-32115

# FURTHER EVIDENCE FOR AN INTERSTELLAR SOURCE OF NIGHTTIME He I 584 Å RADIATION

FRANCESCO PARESCE, STUART BOWYER, AND SHAILENDRA KUMAR*
Department of Astronomy and Space Sciences Laboratory, University of California, Berkeley, California 94720
*Received 1973 November 12; revised 1974 January 2*

## ABSTRACT

The intensity of He I, 584 Å radiation in the night sky was measured from a sounding rocket at altitudes between 160 and 217 km. The results obtained confirm previous measurements of this radiation. The observed intensity and spatial variations can be entirely accounted for by resonance scattering of solar 584 Å radiation from interstellar helium. The best fit to the observations in this case is obtained with an interstellar helium model in which the bulk velocity of the gas is 5 km $s^{-1}$, the temperature is $10^4$ ° K, and the helium number density is 0.032 $cm^{-3}$. Even if a geocoronal scattering component is present, the data can be used to set a lower limit to the temperature of the local interstellar medium of $1.5 \pm 0.5 \times 10^3$ ° K, assuming a normal interstellar helium abundance.

*Subject headings:* interstellar matter — stellar winds

In a previous *Letter* (Paresce, Bowyer, and Kumar 1973), we reported experimental evidence of the existence of diffuse He I, 584 Å radiation in the night sky. This was the result of a measurement made with an extreme-ultraviolet photometer carried on board a Nike Tomahawk sounding rocket launched on 1970 March 10. Since the flux levels observed during this flight were grossly inconsistent with those expected from a geocoronal scattering source (Meier and Weller 1972), we showed that the main features of our observations could be accounted for by backscattering of solar 584 Å radiation from the helium component of the interstellar wind (Paresce, Bowyer, and Kumar 1974). A satisfactory fit to the observational data was obtained with a model having a high helium density and temperature and with a rather large solar line width. Since the observations were made in a region of the sky where the intensities predicted by this model are insensitive to the gas temperature and velocity, these crucial parameters could not be determined accurately.

To explore further the origin of this radiation and to set more stringent limits on the fundamental parameters of the local interstellar medium, an extreme-ultraviolet photometer similar to the one flown on 1970 March 10 was flown to a peak altitude of 217 km on a Black Brant VC sounding rocket launched from White Sands Missile Range, New Mexico, on 1973 February 10 at 0408 UT. At the time of this flight the Sun was in the west, 45° below the horizon. The measurements were made using a 1-cm-diameter funnel-mouthed channel electron-multiplier detector with a thin metallic-film filter consisting of 1600 Å of tin. The combination of photocathode sensitivity and filter transmission restricted the bandpass of the instrument to the range 450–750 Å. Instrumentation of this type has been described in detail by Bowyer *et al.* (1970). Except for the filter material, the detector employed in this study is identical to the one described by Paresce *et al.* (1974), and we refer the reader to that paper for details on the calibration and particle rejection techniques employed. The effective area and solid angle of the device were 0.8 $cm^2$ and 0.012 sr, respectively. The quantum efficiency in counts per photon at 584 Å is 3.2 percent.

The detector was mounted looking out the side of the rocket with a view perpendicular to the spin axis. During the flight the rocket motion was controlled by an attitude control system. The aspect of the detector with respect to the celestial sphere was obtained from star-field photographs taken by a 35-mm camera, star transits in an ultraviolet telescope of the type described by Holberg, Bowyer, and Lampton (1973), and telemetered attitude control performance data. The aspect solution derived in this manner is accurate to $\pm 30'$. In figure 1*a* we show the rocket's altitude in kilometers and the zenith angle of the center of the field of view of the detector as a function of time from launch. The right ascension $\alpha$ and declination $\delta$ of the center of the field of view of the detector throughout the flight are shown in figure 1*b*. As these curves indicate, particular attention on this flight was devoted to observation of the wake of the interstellar wind flow past the Sun where backscattered 584 Å intensities are expected to be most sensitive to temperature and velocity of the interstellar medium (Axford 1972; Paresce and Bowyer 1973).

An altitude- and aspect-dependent signal was recorded by the photometer above approximately 150 km. Below this altitude on both legs of the flight no signal was recorded, indicating that the effect of pinholes was negligible. If it is assumed that there are no other emissions in the detector's bandpass, the resultant He I, 584 Å emission rates in rayleighs (1 rayleigh = $10^6/4\pi$ photons $cm^{-2}$ $s^{-1}$ $sr^{-1}$) may be computed; they are shown in figure 1*c* as a function of time from launch. The peak intensity observed was $6.3 \pm 1$ rayleighs. The error quoted is that associated with the statistical

* Present address: Kitt Peak National Observatory, Tucson, Arizona 85726.

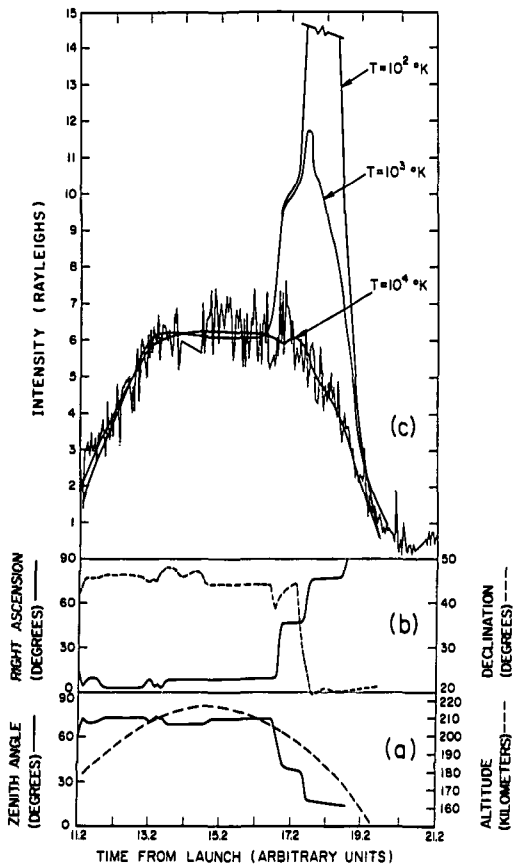



FIG. 1.—(*a*) Rocket altitude in km above the Earth's surface (*dashed line*) and the zenith angle of the center of the photometer's field of view in degrees (*solid line*). (*b*) Right ascension (*solid line*) and declination (*dashed line*) of the center of the photometer's field of view in degrees. (*c*) Observed intensities in Rayleighs as a function of time from launch and predicted intensities computed with interstellar helium model (*smooth solid lines*) using $V = 5$ km s$^{-1}$ and $n(\mathrm{He}) = 0.032$ cm$^{-3}$ for a direction of approach of the interstellar helium of $\alpha = 265°$, $\delta = -20°$.

fluctuations in the measurements only. This result compares favorably with previous measurements of this radiation (Paresce *et al.* 1973, 1974; Riegler and Garmire 1974).

In the following we compare our data with theory by generating the expected 584 Å radiation field due to backscattering from interstellar helium for the geometry applicable at the time of flight and then correcting for absorption within the Earth's geocorona. The procedure employed has been described in detail by Paresce and Bowyer (1973) and Paresce *et al.* (1974). In brief, the intensity of the backscattered 584 Å radiation in photons s$^{-1}$ cm$^{-2}$ sr$^{-1}$ at a frequency $\nu$ and in the view direction $\alpha$, $\delta$ is assumed to be given by

$$I_\nu(\alpha, \delta) = 4\pi^{-1} \int p(\theta) n(s) F_\nu(s) a_\nu(s) ds\,, \qquad (1)$$

where $s$ is the distance along the line of sight measured from the observer.

In this formula $n(s)$ represents the helium number density. The solar 584 Å flux $F_\nu$ is represented by a simple Gaussian which falls off in intensity as the square of the distance from the Sun, and the phase function for scattering, $p(\theta)$, is a function of the angle $\theta$ between the Sun, the scattering particle, and the observer. The scattering coefficient $a_\nu(s)$ is a Gaussian with the central frequency Doppler shifted by the radial velocity of the scatterer with respect to the Sun and with a width which is dependent upon the temperature of the gas.

Assuming the neutral helium to be the result of penetration of the local interstellar medium into the solar system, we have inserted in equation (1) the helium densities derived by Holzer and Axford (1971) for the case $\mu = 0$ (radiation pressure negligible) and the photoionization rate $\beta_0 = 0.62 \times 10^{-7}$ s$^{-1}$. In the wake regions, the formulae developed by Feldman, Lange, and Scherb (1972) specifically for these regions were used. The temperature $T$ and the velocity vector of approach $V$, the helium number density of the local interstellar medium $n$(He), are parameters of this model. The last parameter enters linearly into the intensity calculations, and the results may be scaled accordingly.

The solar 584 Å line width has been measured recently by the University of Colorado group (Cushman, private communication). We have used the observed value of 120 mÅ for the width (FWHM) of the line in our calculations. The intensity emitted at any given point was Doppler shifted by $(1 \pm v/c)$, where $v$ is the velocity of the atoms in the emitting region relative to the Earth. The resulting emission profile was then integrated over frequency, and a region 8 mÅ wide on both sides of the rest frequency was removed to account for scattering by the terrestrial helium, assumed to be at a temperature of 1000° K (Jacchia 1971). The expected intensities obtained in this manner were also corrected for the residual attenuation due to photoabsorption by $O_2$, $N_2$, and O in the line of sight above the rocket. These corrections were made using number densities of the constituents taken from the Jacchia (1971) model atmosphere in combination with the solar 10.7-cm radio data on the day of the flight.

Calculations of the expected interstellar 584 Å emission pattern were carried out using a wide range of values of the variable parameters $T$, $V$, and $n$(He). The ranges of these parameters were chosen on the basis of recent theoretical interpretations (Axford 1972; Thomas 1972; Bertaux and Blamont 1972; Fahr and Lay 1972; Bertaux, Ammar, and Blamont 1972; Wallis 1972; Fahr 1973) of the OGO-5 L$\alpha$ backscatter measurements (Thomas and Krassa 1971; Bertaux and Blamont 1971). According to these interpretations, the temperature of the interstellar medium may lie between 0 and $10^4$ °K, the bulk velocity $V$ of the gas between 5 and 20 km s$^{-1}$, and the hydrogen number density between 0.03 and 0.25 cm$^{-3}$. The direction of approach of the interstellar wind lies in the range $\alpha = 265 \pm 5°$, $\delta = -20 \pm 8°$.

In figure 1*c* we display the results of our calculations. The best fit to the observations is obtained by an inter-

stellar helium model in which $T = 10^4$ ° K, $V = 5$ km s$^{-1}$, and $n$(He) = 0.032 cm$^{-3}$ for the direction of approach $\alpha = 265°$, $\delta = -20°$. The effects of decreasing the temperature from this value are also shown for the cases $T = 10^3$ ° K and $T = 10^2$ ° K for the same values of $V$ and $n$(He). At these lower temperatures the predicted intensities in view directions well above the wake remain almost unchanged, but a prominent maximum is expected when the detector's field of view sweeps across the wake. This peak is a result of gravitational focusing of the interstellar helium atoms in the wake. The absence of such a peak in the data permits us to set a lower limit for the interstellar helium temperature of $5 \pm 1 \times 10^3$ ° K, a temperature which yields predictions that are only marginally consistent with the observations (confidence level = 5% in a $\chi^2$ test) for any value of $V$ and $n$(He) in the ranges specified above. This result remains unchanged if the solar 584 Å line width is varied by ±30 percent, the observational uncertainty. The effect of varying the right ascension of the flow direction in the allowed range is negligible, while a change of the declination from −20° to −28° shifts the position of the expected maxima slightly without affecting the results quoted above. If the declination of the direction of approach of the wind is decreased to the lower limit at −12°, however, the expected intensity maximum falls on the wings of the detector response. For this extreme case the upper limit on $T$ derived from our data is less stringent than for the case of the nominal direction of approach. A gas with a temperature $T = 3 \pm 1 \times 10^3$ ° K yields predictions that are marginally consistent with the data for any value of $V$ and $n$(He) in the specified ranges.

The helium abundance of 0.032 cm$^{-3}$ obtained in our analysis yields a hydrogen abundance of 0.32 cm$^{-3}$ if we assume that the interstellar $n$(He)/$n$(H) ratio is 10 percent (Palmer *et al.* 1969). This value is on the high side but still within the observational uncertainties in the Lα backscatter measurements and consistent with Field, Goldsmith, and Habing's (1969) two-phase theory of the interstellar medium.

A satisfactory fit to the data shown in figure 1*c* can also be obtained with a model in which $T = 10^4$ ° K and $V = 20$ km s$^{-1}$ and $n$(He) = 0.8 cm$^{-3}$. The higher helium density is due to the increasing effect of the Doppler shifts of the scattering atoms (Paresce and Bowyer 1973). Although we cannot *a priori* rule out the possibility that the helium density is indeed this high, we feel this possibility is highly unlikely, and on this basis we reject the model. The present data can thus be used to set an upper limit of $14 \pm 3$ km s$^{-1}$ to the bulk velocity $V$ of the helium atoms, under these assumptions. With this value of the velocity, marginal consistency with the data can be obtained with a value of $n$(He) that is within the uncertainties of the Lα backscatter measurements and the measured $n$(He)/$n$(H) ratio.

Should the geocoronal scattering component be greater than that computed by Meier and Weller (1972), its contribution to the observed intensities may not be negligible. Assuming, however, that the interstellar helium abundance is normal, i.e., $n$(He)/$n$(H) $\simeq 0.1$, the data can still be used to set a significant lower limit of $T = 1.5 \pm 0.5 \times 10^3$ ° K for the temperature of the local interstellar medium. With this temperature and for the extreme value of bulk velocity $V$ of 20 km s$^{-1}$ and $\delta = -12°$, marginal consistency with the data can be obtained. This result rules out truly cold models of the local interstellar medium.

Finally, we cannot exclude the remote possibility that the basic characteristics of the helium component of the interstellar wind such as $V$ and $T$ differ substantially from the ones derived for hydrogen. In this case, clearly, the results derived here could be considerably in error. Only further observations will provide a definitive answer to this important question.

We wish to thank Drs. Michael Lampton and Ray Cruddace for assistance with technical aspects of the experiment and payload integration. We thank Dr. G. Riegler for helpful suggestions and Drs. Bruce Margon and Jay Holberg for their aid in data analysis. We acknowledge the competent launch operations support of the NASA Goddard Space Flight Center Sounding Rocket Division. This work was supported by NASA grant NGR 05-003-450.

## REFERENCES

Axford, W. I. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA SP-308 (Washington: Government Printing Office), p. 609.
Bertaux, J. L., Ammar, A., and Blamont, J. E. 1972, *Space Research*, **12**, 1559.
Bertaux, J. L., and Blamont, J. E. 1971, *Astr. and Ap.*, **11**, 200.
———. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA SP-308 (Washington: Government Printing Office), p. 661.
Bowyer, C. S., Paresce, F., Lampton, M., and Mack, J. 1970, *Planet and Space Sci.*, **18**, 835.
Fahr, H. J. 1973, preprint, University of Bonn.
Fahr, H. J., and Lay, G. 1972, *Mittg der Astr. Gesellschaft*, in press.
Feldman, W. C., Lang, J. J., and Scherb, F. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA Sp-308 (Washington: Government Printing Office), p. 684.
Field, G. B., Goldsmith, D. W., and Habing, H. J. 1969, *Ap. J. (Letters)*, **155**, 149.
Holberg, J., Bowyer, S., and Lampton, M. 1973, *Ap. J. (Letters)*, **180**, 55.
Holzer, T. E., and Axford, W. I. 1971, *J. Geophys. Res.*, **76**, 6965.
Jacchia, L. G. 1971, *Smithsonian Ap. Obs. Spec. Rept.*, No. 332.
Meier, R. R., and Weller, C. S. 1972, *J. Geophys. Res.*, **77**, 1190.
Palmer, P., Zuckerman, B., Penfield, M., Lilley, A. E., and Metzger, P. G. 1969, *Ap. J.*, **156**, 887.
Paresce, F., and Bowyer, S. 1973, *Astr. and Ap.*, **27**, 399.
Paresce, F., Bowyer, S., and Kumar, S. 1973, *Ap. J. (Letters)*, **183**, 87.
———. 1974, *Ap. J.*, in press.
Riegler, G., and Garmire, G. 1974, *J. Geophys. Res.*, in press.
Thomas, G. E. 1972, in *Solar Wind*, ed. C. P. Sonett, P. J. Coleman, Jr., and J. M. Wilcox, NASA SP-308 (Washington: Government Printing Office), p. 668.
Thomas, G. E., and Krassa, R. F. 1971, *Astr. and Ap.*, **11**, 218.
Wallis, M. K. 1972, preprint TRITA-EEP-72-03, Stockholm.

Reprinted from APPLIED OPTICS, Vol. 13, page 575, March 1974
Copyright 1974 by the Optical Society of America and reprinted by permission of the copyright owner

# Extreme Ultraviolet Spectrometer for Space Research

Shailendra Kumar, Francesco Paresce, Stuart Bowyer, and Michael Lampton

A74-23420

A normal incidence spectrometer for use in the wavelength region from 200 Å to 1270 Å has been developed. The design and calibration of the instrument are described in detail. The spectrometer can be employed to detect extreme uv radiation at a minimum flux level of 1 rayleigh with a spectral resolution of 40 Å and a spatial resolution of 6°. Data on the extreme uv night sky spectrum between 780 Å and 1270 Å obtained with this instrument on a recent rocket flight are presented.

## Introduction

Most of the strong resonance lines and bands of the atomic, ionic, and molecular constituents of the earth's upper atmosphere and interplanetary medium lie in the far (1050–2000-Å) and extreme (200–1050-Å) ultraviolet. Since this part of the spectrum is unobservable from the ground, studies of these wavelength regions must be carried out from rockets or satellites. Because of the difficulties inherent in detecting radiation in the extreme ultraviolet (EUV) at low intensities in space, the existing data on the nighttime EUV flux are limited to a small number of measurements made with broadband ($\Delta\lambda \simeq$ 300-Å) photometers.[1-4] The interpretation of these data is necessarily restricted by the assumptions concerning the wavelengths of the radiation being observed.

More information exists on the EUV dayglow spectrum. In addition to a number of broadband photometric measurements,[3,5-7] Carruthers and Page[8] have obtained some spectrographic data in this wavelength region. Their instrumentation was not sufficiently sensitive, however, to detect the fainter lines expected in the daytime airglow and to observe below 500 Å, where strong lines of neon and helium are expected.

In order to study the entire EUV spectrum with sufficient sensitivity to detect weak lines, we have developed a versatile spectrometer that provides high spectral resolution ($\simeq$10–40 Å) and moderate spatial resolution ($\simeq$6°).

## Spectrometer Optical Design

Since most of the airglow emissions in the EUV are relatively faint, the primary requisite of any spectrometer designed for these wavelengths is high sensitivity. Most spectrometer designs are limited by the fact that the incident radiation must undergo at least two near normal reflections before arriving at the detector. Because the reflectances of materials at EUV wavelengths are very small, all these designs have low sensitivities. We have considered three types of spectrometer mounts that need only one reflection of the incident beam as to their suitability for EUV airglow measurements: a grazing incidence Rowland circle mount; a normal incidence Wadsworth mount; and a normal incidence Rowland circle mount. The most suitable design for our purposes was selected from these three options by requiring that the instrument have the following characteristics: highest possible sensitivity to diffuse radiation; moderate spatial and spectral resolution; a diffracted image of small size; and mechanical simplicity.

The grazing incidence mount has a number of attractive features for use in the EUV (small size, high sensitivity), but it has two major shortcomings. First, the image is highly astigmatic. At an angle of incidence of 82° with a concave grating ruled at 2400 lines/mm and a groove length of 3 cm, the image of a point at the entrance slit in 600-Å light is a line 10 cm high in the focal plane. Consequently a 2-cm long entrance slit is imaged into a line 14 cm high. This creates a problem in arranging an EUV detector at the exit slit that is sufficiently large to collect all the dispersed light. Second, because of the dimensions and proximity of the entrance slit to the grating, this grazing incidence spectrometer has a large field of view making data interpretation and calibration difficult. Although the field of view can be reduced by using a small slit and masking the grating or by decreasing the angle of incidence, both of these modifications result in a loss of sensitivity. An additional drawback in the grazing incidence design is the mechanical complexity of the scanning mechanism.

The Wadsworth mount utilizes a parallel beam of light and is best suited to those applications in

All authors were with the Department of Astronomy and Space Sciences Laboratory, University of California, Berkeley, California 94720 when this work was done; S. Kumar is now at Kitt Peak National Observatory, Tucson, Arizona 85726.

Received 30 July 1973.

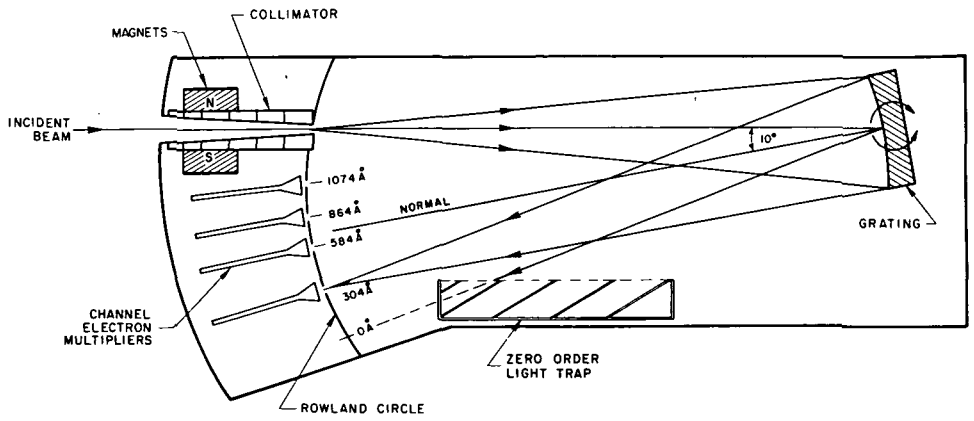



Fig. 1. Optical layout of the EUV spectrometer.

which the incident light is parallel or near parallel. Two techniques can be employed if the beam to be measured is not parallel as will be the case for a diffuse airglow or interplanetary background. One approach is to use a collimating mirror with a slit at the focal point of the mirror; the other is to use a passive collimation system to render the incoming beam parallel. The first approach results in a two-reflection instrument and offers no improvement over standard mounts. The second technique requires the incident radiation in the dispersion direction be collimated to a few min of arc for moderate spectral resolutions. The resulting loss in solid angle reduces the sensitivity to diffuse radiation and must be compensated for by using a large entrance aperture and a large field of view in the direction perpendicular to the dispersion axis. The resulting instrument is relatively difficult to baffle against sunlight.

The normal incidence Rowland circle mount is free from astigmatism, simple to baffle, provides a convenient field of view ($\sim$6°) with moderate spectral resolutions (10–40 Å), and the scanning mechanism is very simple. For these reasons and because of its inherently high sensitivity and mechanical simplicity, we find this design more suitable for measurements of a source of diffuse radiation. We describe in this paper the spectrometer that we have constructed using this mount.

The optical layout of the spectrometer is shown schematically in Fig. 1, and its characteristics are given in Table I. The incident light first passes through a baffle to eliminate off-axis radiation. The entrance slit we employed is 4 mm wide and 2 cm high. The light impinges on the grating at an angle of incidence of $\sim$10°. The diffracted EUV light is detected at four wavelengths simultaneously by means of detectors placed behind four exit slits (20 mm × 4 mm) in the image plane. We employed a platinum-coated concave grating blazed at 304 Å. This grating had a 400.7-mm radius and was ruled at 2400 lines/mm. With this grating a theoretical resolution of $\sim$0.006 Å at 584 Å can be obtained, because we are using wide slits to provide high sensitivity, the resolution is determined by the slit width itself. Since the dispersion of this instrument is 10 Å/mm, the spectral resolution for 4 mm slits is about 40 Å. It is important to note that better spectral resolutions can be achieved at the expense of sensitivity in this design by reducing the size of the entrance and exit slits. As an example, 1-mm slits would provide 10-Å resolution, which would be desirable in a daytime measurement where large intensities of the spectral lines are expected, and therefore a large throughput is not required.

The first-order inside spectrum is used because this arrangement results in considerable saving in space as compared to the outside order. For the grating selected this configuration also insures maximum intensity of the diffracted beam. For angles of incidence and diffraction less than 10°, the dispersion and the grating reflectance are essentially independent of angle, making the instrument easy to calibrate. The astigmatism of the instrument is negligible. For example, at 500 Å the image length of the slit is increased by only about 1/30 of the groove length. For 3-cm long rulings, this means that image length of a 20-mm long entrance slit is increased by only 1 mm. With such a small entrance slit it is relatively easy to baffle the spectrometer against off-axis sunlight.

**Table I. EUV Airglow Spectrometer Characteristics**

| | |
|---|---|
| Spectral range | 220–1270 Å |
| Spectral resolution | 41.6 Å |
| Concave grating: | |
| Focal length | 400.7 mm |
| Grating constant ($1/d$) | 2400 lines/mm |
| Ruled area | 30 × 38 mm |
| Blaze wavelength | 304 Å in first order |
| Reflective coating | Platinum |
| Slit width | 4 mm |
| Slit length | 20 mm |
| Field of view | 6 × 7° |
| Geometry factor | $5.4 \times 10^{-3}$ cm² sr |
| Angle of incidence | 10° |
| Angular dispersion | 72.7 Å/degrees |
| Linear dispersion | 10.4 Å/mm |
| Wavelength scan | 10 Å/sec |
| Scan period | 28 sec |
| Accumulation time (per line per scan) | $\sim$4 sec |

**Table II. EUV Spectrometer Scan**

| | Filter | Spectral range Å | Expected airglow emissions Å |
|---|---|---|---|
| Detector 1 | 1000 Å Al | 220–500 | He II 304 |
| Detector 2 | 1000 Å Al | 500–780 | He I 584 |
| | | | Ne I 736 |
| Detector 3 | 1560 Å In | 780–1060 | 0 II 834 |
| | | | $N_2$-Birge Hopfield bands |
| | | | H I 1025 |
| Detector 4 | none | 990–1270 | H I 1216 |

### Mechanical Assembly

The grating drive mechanism and the detector section are built as separate housings that are mounted on a rigid aluminum base plate. The spectrometer can be operated in two alternate modes. The first is a scanning mode in which the grating rotates so as to sweep the EUV spectrum across the detector entrance slits. For this mode the grating is held on a shaft that rotates in ball bearings mounted in the top and bottom of a support casting. The grating drive consists of a continuous motor/cam follower system that provides a 24-sec scan in the forward direction and a 1-sec flyback. A shaft encoder with analog output is provided to determine the position of the grating at any given instant.

Four detectors units are used in the image plane so that four different parts of the spectrum can be scanned simultaneously. Each detector scans 280 Å of the spectrum corresponding to about a 2° rotation of the grating. In this arrangement, each spectral line is accumulated in the respective detector unit for 4 sec/scan at 40-Å resolution. Since this spectrometer was designed to explore the entire EUV spectrum, the positions of the detectors were chosen to cover as large a part of the spectrum as possible. The spectral ranges covered and the expected airglow emission in these ranges are given in Table II.

In an alternate fixed grating mode of operation, the spectrometer can be used to monitor four different lines of the spectrum simultaneously. For example, a desirable set of observations might be the simultaneous measurement of the 1216-Å and 1025-Å lines of hydrogen and the 304-Å and 584-Å lines of helium.

Major sources of background in this kind of measurement are scattered EUV radiation and charged particles. Several precautions are taken to reduce scattering of EUV radiation inside the spectrometer housing. First, the entrance slit was baffled so that the only direct path for the light is from the slit to the grating itself. All the baffles and other surfaces where secondary scattering could occur are coated with gold black. A gold black baffle system is also placed in a position to trap the zero order light. Hydrogen Lyman $\alpha$ airglow emission is usually the largest source of scattered radiation, hence the detectors are equipped with metallic bandpass filters that discriminate against this radiation. Aluminum filters are used for the 220–780-Å range and an indium filter is used for the 780–1060-Å range. The fourth detector for the 990–1170-Å range is used without any filter.

To shield against charged particle background, two magnets are attached to the collimator baffle; these provide a field of about 600 G inside the collimator. This assembly is capable of rejecting electrons and protons with energies up to tens of kilovolts.

The entire spectrometer weighs 6.84 kg. A photograph of the instrument without its cover is shown in Fig. 2.

### Detector and Electronics Assembly

The photon detectors employed are spiraltron channel electron multipliers with 1-cm diam funnels. These detectors have high quantum efficiency in the EUV that is coupled with a low intrinsic background.[9] Two detectors that are tied together electronically as a single unit are used at each of the four exit slits. The four units are electrically shielded from each other to avoid possible cross talk. Pump-

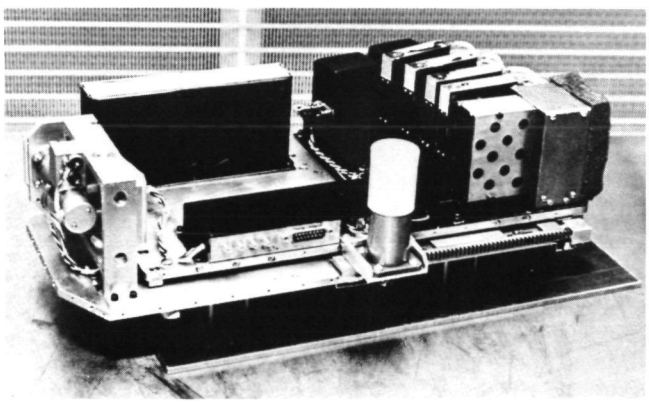


Fig. 2. The instrument without its protective cover. In this view the grating is on the left, and the collimated entrance slit and detectors are on the right.

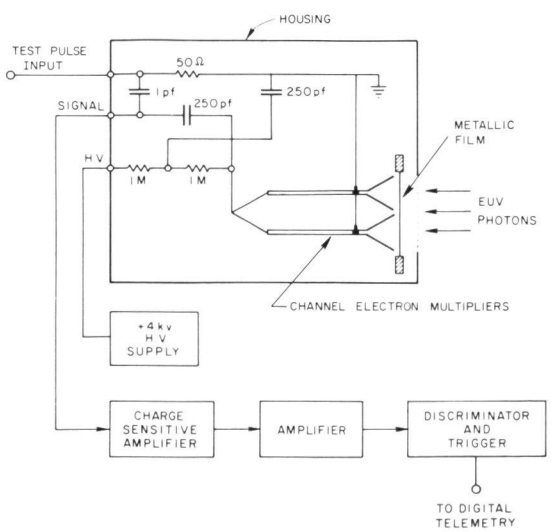



Fig. 3. Schematic of the detector unit and associated electronics.

out ports are provided through the shielding, and since these ports could in principle allow thermal electrons to reach the detectors, a −29 V grid is mounted on each of these ports. Each of the four units is equipped with its own HV power supply to avoid a possible HV breakdown in one detector from affecting the other detectors.

The signal pulse from each unit is fed to a charge integrating preamp. The output of the preamp is then fed through appropriate amplifiers to a level discriminator/trigger unit. The amplifier gain is selected to allow at least a factor of 100 degradation in detector gain without loss of signal. The discriminator/trigger output is then fed to the telemetry system. A schematic of one of the detector units and its associated electronics is shown in Fig. 3.

The total power consumption of the spectrometer is 6.2 W (~220 MA at +28 V).

### Instrument Calibration

The sensitivity of the spectrometer in counts $sec^{-1}$ $rayleigh^{-1}$ at a given wavelength is given by the relation

$$N/I = (10^6/4\pi) \times R\eta ETGC_1C_2, \quad (1)$$

where $N$ is the measured count rate in $sec^{-1}$, $I$ is the intensity of the incident radiation in rayleighs, $R$ is the reflectance of the grating, and $\eta$ is the grating blaze efficiency, i.e., the fraction of the diffracted radiation that is diffracted into the inside first order. The quantum efficiency of the spiraltron, the transmission of the metal filter, and the geometry factor in $cm^2$ sr are indicated by $E$, $T$, and $G$, respectively. The normalization factors for the variation of $R\eta$ over the ruled grating area and for the variation of quantum efficiency along the exit slit length are given by $C_1$ and $C_2$, respectively. The geometry factor can be further expressed as $G = A_gA_s/D^2$, where $A_s$ is the area of the entrance slit, $A_g$ is the exposed area of the grating, and $D$ is the distance between the entrance slit and the grating, which for this mount is the radius of curvature of the grating.

The response of the spectrometer was estimated by using nominal data on the channel electron multiplier quantum efficiency[10] and typical data on the blaze efficiency $\eta$ for this type of grating from Samson.[11] The normal incidence reflectance $R$ of platinum and the transmission of the aluminum and indium filters at EUV wavelengths were also taken from Samson.[11] The variation of these parameters with wavelength is shown in Fig. 4.

An absolute calibration of the spectrometer was carried out at several EUV wavelengths. A high intensity continuous discharge hollow cathode source[12] and a Seya Namioka monochromator were employed to generate EUV radiation in the laboratory. The quantum efficiency of the detectors was determined by reference to a gold diode that was used as an absolute intensity standard.

A channel electron multiplier calibrated against the gold diode was used as a secondary standard. An on-axis collimated beam (spread ~1° FWHM) was used to obtain the product $R\eta ET$, which is given by

$$R\eta ET = (N_s/N_c)E_c, \quad (2)$$

where $N_s$ is the count rate in a given spectrometer channel, $N_c$ is the count rate in the secondary standard, and $E_c$ is the absolute quantum efficiency of the secondary standard.

The normalization factor $C_1$ was obtained by fixing the photon beam while the spectrometer was rotated in the dispersion plane about a vertical axis passing through the entrance slit. The normalization factor is given by the relation,

$$C_1 = \frac{1}{\theta_0}\int_{-\theta_0/2}^{+\theta_0/2} \frac{N(\theta)}{N(0)}d\theta, \quad (3)$$

where $\theta$ is measured from the spectrometer axis which is defined as the axis passing through the center of the grating and center of the entrance slit; $\theta_0$ is the full width of the field of view in the dispersion direction, $N(\theta)$ is the count rate measured when the photon beam is incident at angle $\theta$ and $N(0)$ is the count rate measured at $\theta = 0°$.

The normalization factor $C_2$ was obtained by keeping the photon beam parallel to the spectrometer axis while scanning along the length of the slit. Here

$$C_2 = \frac{1}{l}\int_{-l/2}^{+l/2} \frac{N(x)}{N(0)}dx. \quad (4)$$

Where $l$ is the length of the slit, $N(x)$ is the count

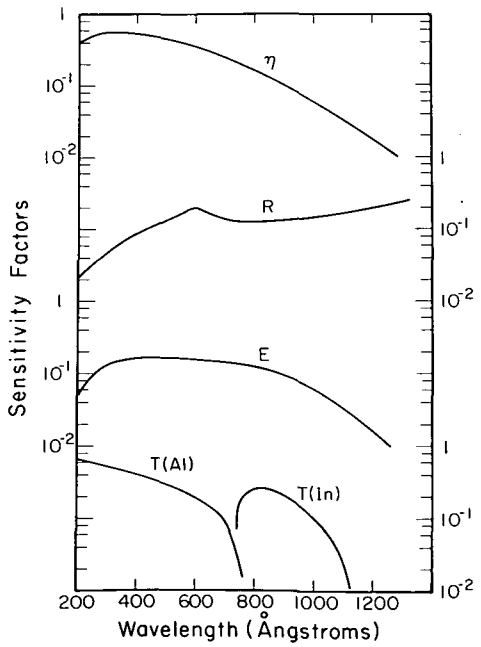



Fig. 4. The sensitivity factors of the spectrometer: $\eta$, grating blaze efficiency; $R$, grating reflectance; $E$, spiraltron quantum efficiency; $T$(Al), transmission of aluminum filter; $T$(In), transmission of indium filter.

rate measured when the photon beam is incident at the position $x$ along the slit, and $N(0)$ is the count rate measured when the photon beam is incident at the center of the slit.

The absolute calibration of the detectors is shown by the data points in Fig. 5. The solid lines in this figure show the expected sensitivity profiles that were calculated by using the nominal data shown in Fig. 4. The nominal response calculated from these data may vary from the actual response because of differences between the actual values of the efficiency of the grating and the channel electron multipliers and the nominal values assumed in the calculations.

The scattering efficiency of $H$ Lyman $\alpha$ radiation into detectors 1, 2, and 3 was measured and was found to be less than $10^{-4}$. The over-all background count rates of each of the detector channels from all sources are less than 0.5 counts $sec^{-1}$; thus the minimum detectable flux should be of the order of 1 rayleigh for an integration time of 4 sec.

A number of laboratory scans were obtained to calibrate the positions of the spectral lines. The zero order light from the Seya Namioka monochromator was used as the incident radiation to obtain the scans from the rocket spectrometer. A typical scan obtained by using helium gas in the hollow cathode discharge source is given in Fig. 6. The upper four curves are the scans obtained by the four detectors of the rocket spectrometer. The respective spectral ranges are marked in Angstroms on each scan. For comparison, the corresponding spectral scans obtained by the Seya Namioka monochromator (resolution $\sim$1 Å) are shown in the lower four curves in this figure. At wavelengths >800 Å, many lines are present in the Seya Namioka scan that are smeared out in the spectrometer scans due to the 40-Å resolution of the instrument.

The strong resonance line of neutral helium at 584 Å can be easily seen in the 500–780-Å scan. The line appears asymmetric because the observed spectral feature is the superposition of the 584-Å and 537-Å lines of helium. The 304-Å resonance line of $He^+$ is recorded in the 220–500-Å detector. The relatively large count rate observed with this detector at wavelengths $\geq$400 Å is due to the scattering of the extremely intense 584-Å line.

These scans show that the 40-Å resolution of the rocket spectrometer is sufficient to identify the major emission lines of helium. Similar scans were obtained for other atmospheric gases, namely hydrogen, oxygen, argon, neon, etc., and these scans showed that this spectrometer will be able to identify EUV airglow emissions of these gases if their intensities are greater than $\simeq$1 rayleigh and if their lines are sufficiently separated. If a large number of lines are present in the EUV airglow spectrum, smaller slit widths with a resulting higher resolution will be required for line identifications.

## Flight Performance

The spectrometer was flown aboard a Black Brant V-C rocket launched on 9 February 1973 at 04 08 UT (21:08 LT) from White Sands Missile Range in New Mexico. The rocket reached a peak altitude of 210 km. The instrument was turned on at 170 km and spent a total of $\simeq$220 sec viewing large area of the sky between 180° and 310° azimuth and between 10° and 70° zenith angle. The count rates from detectors 1, 2, and 3 were very low as expected. Detector 4 had a large background of undetermined origin throughout the flight. All the scans during the flight were added together to obtain the highest SNR. The flight data for the detector channels scanning the 990–1270-Å and the 780–1060-Å bands are shown in Figs. 7 and 8. In Fig. 7 the intense airglow Lyman alpha line stands out clearly above the

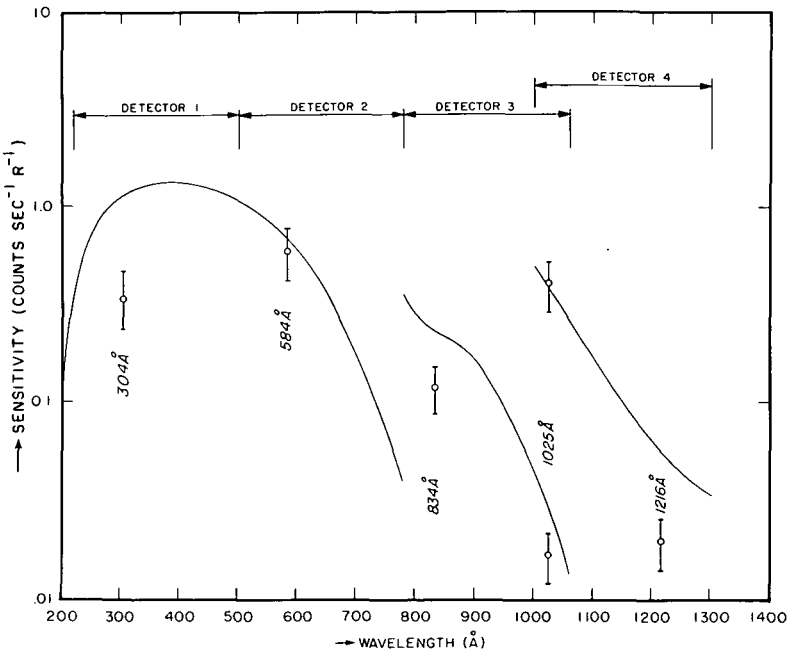



Fig. 5. Absolute calibration of the spectrometer. The open circles show the sensitivity as determined in our laboratory; the solid lines show the sensitivity calculated on the basis of nominal data on the optical components.

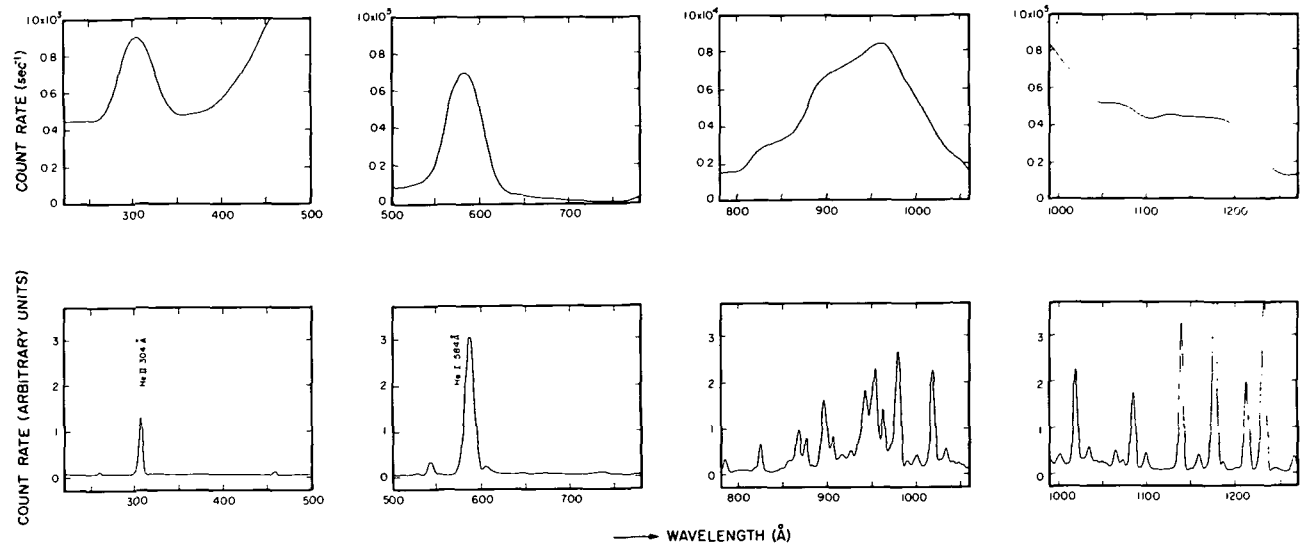



Fig. 6. Response of the spectrometer to a laboratory source. The bottom row of curves shows the input spectrum as observed with a Seya Namioka spectrometer with resolution of 1 Å. The top row of curves shows the response of the spectrometer. The 40-Å resolution of the instrument is clearly insufficient to resolve the large number of lines that are present in the laboratory source at wavelengths longward of 800 Å.

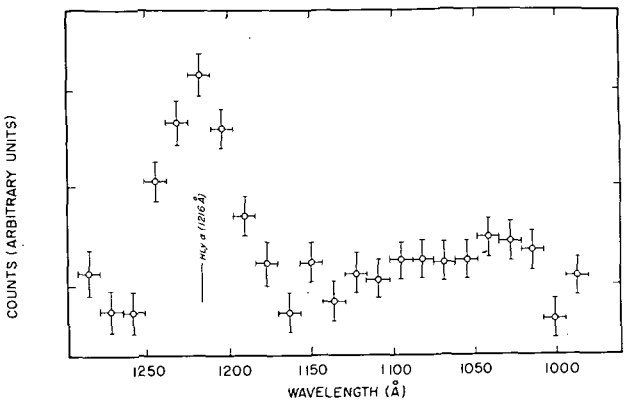



Fig. 7. Flight data in the 990–1270-Å band. The hydrogen Lyman alpha line is clearly present in these data.

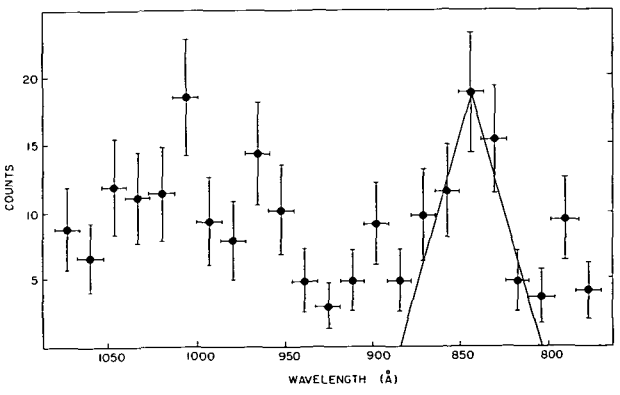



Fig. 8. Flight data in the 780–1060-Å band.

background. As expected its full width at half maximum is 40 Å. There is evidence of a spectral line near 840 Å that may be identified as the 834-Å resonance emission of $0^+$. A number of lines could contribute to the signal observed at 920–1050-Å wavelengths. The primary candidates are HI 1025 Å, OI 990°, OI 1025 Å, and Birge Hopfield bands of $N_2(b^1\Pi_u - X^1\Sigma_g^+)$.

The airglow results from these data will be reported in a separate publication.[13]

The authors thank J. Buckley for a number of useful discussions and Ann Rotramel for editorial assistance. This research was supported in part by NASA grant 05-003-450.

## References

1. C. S. Bowyer, P. M. Livingston, and R. D. Price, J. Geophys. Res. **73,** 1107 (1968).
2. J. M. Young, C. S. Weller, C. Y. Johnson, and J. C. Holmes, J. Geophys. Res. **76,** 3710 (1971).
3. T. Ogawa and T. Tohmatsu, J. Geophys. Res. **76,** 6136 (1971).
4. F. Paresce, S. Bowyer, and S. Kumar, J. Geophys. Res. **78,** 71 (1973).
5. T. M. Donahue and J. B. Kumer, J. Geophys. Res. **76,** 1145 (1971).
6. S. Kumar, S. Bowyer, and M. Lampton, J. Geophys. Res. **78,** 1107 (1973).
7. R. W. Carlson and D. L. Judge, Planet. Space Sci. **21,** 879 (1973).
8. G. Carruthers and T. Page, Science **177,** 788 (1972).
9. W. G. Wolber, B. D. Klettke, and P. W. Graves, Rev. Sci. Instrum. **41,** 724 (1970).
10. M. C. Johnson, Rev. Sci. Instrum. **40,** 311 (1969).
11. J. A. R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy* (Wiley, New York, 1967).
12. F. Paresce, S. Kumar, and C. S. Bowyer, Appl. Opt. **10,** 1904 (1971).
13. S. Kumar, S. Bowyer, and F. Paresce, submitted to J. Geophys. Res. (September 1973).

which the current system is enclosed in the moon and is not subject to attenuation by an insulating surface layer. The driving function is an enhanced solar magnetic field convected outward from a rapidly spinning sun by a dense solar wind corresponding to a solar T Tauri phase.

P24

Román Alvarez
Engineering Geoscience
University of California
Berkeley, Calif. 94720

(Now at Instituto de Geofisica, Univ. Nacional de Mexico, Mexico 20, D.F.)

Lunar Fines 74241,2: Low-Frequency Dielectric Variations with Density and Temperature. The dielectric variations of lunar sample 74241,2 were obtained in the frequency range of 30 to $10^5$ Hz, in ~$10^{-9}$ torr vacuums, as a function of temperature and density. Temperatures of 100°, 298°, and 373°K, and packing densities of 1.38 ± .05 and 1.61 ± .05 g/cm$^3$ were measured. The sample was handled in a 0% relative humidity atmosphere of $N_2$ prior and after measurements in the high vacuum. Initial control measurements, at 298°K and 2.0 torr of $N_2$ showed a low-frequency dispersion which disappeared after evacuation, suggesting that the sample was slightly contaminated or that volatiles accumulated in it during storage and transport. The dielectric constant ($\kappa'$) value for the $\rho$= 1.38 g/cm$^3$ sample at room temperature and 100 KHz was 2.21, increasing to 2.33 at 30 Hz. The $\rho$= 1.61 g/cm$^3$ sample yielded $\kappa'$ values of 2.38 at 100KHz and 2.54 at 30 Hz. The dielectric losses ($\tan\delta$) remain constant with density at approximately .013. Variations within ± .10 of the 298°K values of $\kappa'$ were obtained for the 373° and 100°K responses, the larger values corresponding to the higher temperatures. The $\tan\delta$ values showed small variations with temperature at frequencies below $10^3$ Hz. These observations support the suggestion that density variations prevail over temperature variations as the controlling agent of dielectric properties in the lunar regolith.

P25

Richard A. Simpson
Ctr for Radar Astronomy,
Stanford, Calif. 94305

Subradar Point Variation and Its Effect on Monostatic Radar Studies of the Moon. Recent monostatic radar studies of the moon at 1290 MHz have indicated that surface roughness variations at different subradar points can be measured from power vs time curves of the returning short pulse echoes. Estimates of rms unidirectional slope on horizontal scale lengths of 50–100 m are near 5° in Sinus Medii and the Schröter region but rise to over 7° in the highlands around Hipparchus.

P26

J. E. Keith
R. S. Clark
Geochemistry Branch, NASA
Johnson Space Center,
Houston, TX 77058

The Saturation Activity of $^{26}Al$ in Lunar Samples as a Function of Chemical Composition. The saturated specific activity is shown to be a function of the silica, alumina and magnesia contents as well as the reciprocal of the cube root of the sample weight, by linear multivariate regression. The latter term is introduced to remove the dependence of the specific activity on the surface to volume ratio of the sample. Since only unaltered lunar samples are considered, the resulting expression is only applicable to such samples. For those 29 samples whose activities and compositions are known, 21 are shown to be saturated and all those in the set previously thought to be unsaturated - 12034, 12064, 14301 and 14305,38 - are shown to be unsaturated by internal consistency arguments. In addition, 15265, 15415, 10003 and marginally, 60316 are shown to be unsaturated. The expression representing the specific activity of the saturated set is:

$$^{26}Al(dpm/kg) = 0.652[SiO_2] + 2.50[Al_2O_3] + 0.560[Al_2O_3]\,W^{-1/3} + 2.28[MgO]\,W^{-1/3}$$

where $[M_xO_y]$ is the weight percent oxide as percent and W is the sample weight in kilograms.

P27

J. P. Bibring*
A. L. Burlingame†
J. Chaumont*
M. Maurette*
F. C. Walls†
P. C. Wszolek†

*Laboratoire Rene Bernas, 91406 Orsay, France

†Space Sciences Lab., University of California, Berkeley, Calif. 94720

The Ion Implantation Chemistry of Carbon in Metals and Mineral Grains and its Implications for Cosmochemistry. Various targets including metals and minerals found in the lunar regolith have been artificially subjected to heavy doses of solar wind type ions: $^{13}C^+$, $^{14}N^+$, and $D_2^+$. The physics of ion implantation was carefully monitored by observing the irradiated targets with a high voltage electron microscope. Molecules possibly synthesized during these experiments were sought with a high resolution ionic analyzer and the high resolution mass spectrometry techniques already applied to the study of lunar carbon compounds. Several synthesized molecular species were detected: $^{13}CD_4$, $^{13}C_2D_4$, $^{13}C_2D_6$, and $^{13}C_3D_6$. The synthesis of calcium and/or sodium carbides was also inferred from the detection of $^{13}C_2H_2$ as the reaction product of irradiated plagioclase with hydrofluoric acid (HF). Additional molecules were released by heating the plagioclase targets during pyrolysis experiments: HOD (400-600°C); $H^{13}CN$ (800-900°C); $^{13}CO$ (700-900°C) plus 1100-1200°C); and $^{13}CO_2$ (800-900°C plus 1000-1200°C). These results have reproduced some of the features of lunar carbon chemistry and also have important implications for the carbon chemistry of dust grains in interstellar clouds and the early solar nebula.

P28

M. J. Drake
Dept. of Planetary Sciences
Univ. of Arizona, Tucson, Arizona 85721
G. J. Taylor
Dept. of Earth Sciences
Washington University,
St. Louis, Missouri 63136
G. G. Goles
Center for Volcanology
Univ. of Oregon, Eugene, Oregon 97403

Geochemistry of petrologically-characterized lithic fragments from Apollo 16. We discuss the geochemistry of twenty eight petrologically-characterized 2-4 mm lithic fragments separated from five Apollo 16 soil samples. The fragments are subdivided on the basis of petrologic criteria into four subgroups: the ANT suite, K-rich residual mesostasis-bearing rocks, poikiloblastic rocks, and (spinel)-troctolites. Feldspathic basalts analyzed in this study appear to be impact melts of pre-existing rocks. Light-matrix breccias were not represented among the samples analyzed. It is shown that REE concentrations in ANT rocks increase sympathetically with increasing abundance of phases other than plagioclase. In K-rich residual mesostasis-bearing rocks REE concentrations are directly related to the modal percentage of the residual phases. The provenance of the Cayley Formation is discussed and it is concluded that the light plains-forming deposits are unlikely to originate as Orientale ejecta. Model calculations of possible parent liquids for ANT rocks indicate that initial REE concentrations exceeded chondritic values by factors of 5-8.

P62

R. W. Perkins
L. A. Rancitelli
W. D. Felix
N. A. Wogman
Battelle
Pacific Northwest
Laboratories
Richland, Washington
99352
Sponsor David L. Schreiber

Lunar Surface and Solar Process Characterization From Apollo 16 Sample Studies. The cosmogenic radionuclides 22-Na, 26-Al, 46-Sc, 54-Mn, 56-Co, 48-V and 60-Co and the primordial radionuclides K, U and Th were measured in 12 soils, 3 crystalline rocks and 11 breccias. The unusual chemistry of that area resulted in higher 26-Al and lower 22-Na concentrations from low energy solar protons than at previous Apollo sites. The concentrations of 48-V, 46-Sc, 54-Mn and 56-Co were the lowest observed in any Apollo samples. Friable soil clods collected in rake samples were in general undersaturated in 26-Al. Average lunar surface ages of soil clods ranged from $< 10^5$ years for 65511,9 to $\sim 10^6$ years for 63501,4. Skim soil sample (69941,1) and the soil sample (69921,1) taken directly below it showed essentially identical 26-Al to 22-Na ratios, indicating substantial mixing to a few cms. The ranges in K, Th, and U in soils were 0.07% to 0.12%, 1.76 to 2.74 ppm, and 0.41 to 0.74 ppm, respectively. The rocks showed much wider ranges: K, < 0.009 to 0.195%; Th, 0.14 to 3.98 ppm; and U, 0.038 to 1.18 ppm, while chips (69935 and 69955) from a large boulder showed a 20-fold variation in their K, U, and Th content.

P63

Patricia C. Wszolek
B. R. Simoneit
P. T. Holland*
A. L. Burlingame
Space Sciences Lab., Univ. of Calif., Berkeley, Calif. 94720
(*Now at Ruakura Soil Research Station, Hamilton, New Zealand)

Meteoritic Contribution to Lunar Carbon Chemistry. To determine the amount of carbon added to lunar soil by meteorites, we have separated fractions enriched in magnetic grains from two samples of mature Apollo 14 soils, 14003 and 14163, and from a Plum Crater rim soil, 61500. Carbon chemistry of the magnetic materials was compared with that of nonmagnetic materials by high resolution mass spectrometric analysis of the products of DF acid dissolution and vacuum pyrolysis to 1400°C. Our data indicate that there has been a direct input of carbon into these soils from meteorites. In the Apollo 14 soils studied, a maximum of 2% of the carbide and 2 to 4 ppm total carbon appears to derive from micrometeorites. Soil 61500 probably includes ejecta material from North Ray, South Ray, Flag, and Plum Craters and probably retains some components of one or more of the projectiles which formed these craters. Our preliminary estimate of meteoritic contribution to this soil is about 33% of the carbide and 2.1 ppm total carbon. We have also examined subsurface soil 61221,7 from Plum Crater. This sample was found to have had short exposure time on the lunar surface, but it is rich in volatilizable matter such as $H_2O$, CO, $CO_2$ and smaller amounts of hydrocarbons, HCN, oxygenated species ($C_nH_{2n-z}O$ for n = 1 to 4 and z = -2 to 4), and nitrogenous species ($C_nH_{2n-z}N$ for n = 1 to 3 and z = 0 to 4). These volatiles are thought to derive from a projectile, perhaps a carbonaceous chondrite or a comet.

P64

Fouad Tera
G. J. Wasserburg
Lunatic Asylum, Division of Geological and Planetary Sciences, California Institute of Technology
Pasadena, California 91109

Lead Isotopic Evidence on Lunar Evolution. U-Pb internal isochrons for a number of lunar rocks were determined through newly developed micro-analytical techniques with a total blank of $\sim 3 \times 10^{-10}$ gm of Pb. Crystallization ages thus obtained are in reasonable agreement with ages obtained by other methods. The presence of radiogenic initial Pb was proven to be common to most lunar rocks including mare basalts. All highland soils from Apollo 16 and Luna 20 missions are characterized by high $^{207}Pb/^{206}Pb$ values ($\sim 0.8$). This is attributed to a component in the soils containing radiogenic Pb produced in the first 0.6 AE of the moon's history and seems to be correlated with the initial Pb found in the rocks. We suggest that a major lunar cataclysm took place $\sim 3.95$ AE ago. This event may be associated with the formation of a major basin, possibly Imbrium. This process excavated ancient materials rich in U and Th either from a layer which originally lay beneath an outer crustal layer, which included ancient anorthosites, or from patches which cross-cut the ancient crust. Pb with isotopic composition similar to that of the initial Pb common to the rocks from Apollo 14, and 16 sites and rock 15415 was widely dispersed over the lunar surface by this impact. The isotopic composition of fragments of ancient lunar crustal materials in the regolith must have been altered by this process.

P65

Robin Brett
R. C. Gooley
NASA Manned Spacecraft Center, Houston, TX 77058
Eric Dowty
Martin Prinz
Klaus Keil
Dept. of Geology & Inst. of Meteoritics, University of New Mexico, Albuquerque, NM 87106

Oxide Minerals in Lithic Fragments from Luna 20 Fines. 176 oxide mineral grains in the Luna 20 sample were analyzed by electron microprobe. The three most abundant oxide minerals are spinel, ilmenite, chromite, with subordinate ulvöspinel, armalcolite, zirkelite, baddeleyite, and an unidentified brick-red $TiO_2$-rich phase. Metallic (approx. $Fe_{99}Ni_1$) rods occur in some plagioclase grains in up to 4 crystallographically controlled orientations; these may have formed by sub-solidus reduction of FeO in plagioclase. Ilmenite of mare origin in the Luna 20 sample is Mg-poor and Zr-rich compared to non-mare ilmenite; these elements may therefore be useful in determining the origin of ilmenite grains. Spinels are predominantly spinel-hercynite ss. with Mg/Mg+Fe greater than in spinel from Apollo 14 rocks. Spinels occur exclusively in spinel troctolite fragments. Phase equilibria considerations suggest that if spinel troctolite assemblages are cumulates, they crystallized from a melt high in alumina; a likely candidate is the high alumina basalt of Prinz et al. (1973) since it plots very close to the peritectic piercing point, spinel-plagioclase-olivine in the system $Fe_2SiO_4$-$Mg_2SiO_4$-$CaAl_2Si_2O_8$-$SiO_2$ of Walker et al. (1972). Other lunar compositions of possible primary origin either could not produce spinel troctolite or would require very extensive plagioclase removal at low pressures.

P66

C.E. Nehru*
Martin Prinz
Eric Dowty
Klaus Keil
Dept. of Geology, Univ. of New Mexico, Albuquerque, 87106
(*Permanent address: Brooklyn College, Brooklyn, N. Y. 11210)

Spinels and ilmenites in Apollo 15 Igneous Rake Samples. We obtained 502 spinel and 163 ilmenite analyses from 33 Apollo 15 rake samples of igneous origin using the electron microprobe. Textural relationships were also studied. Mare basalts: Seven olivine basalts and 12 olivine microgabbros have spinels in the chromite-ulvöspinel series with ulvöspinels sometimes mantling chromites; some show complete solid solution, although intermediate compositions are sparse. Two olivine-rich gabbros, similar in composition to green glass, seem to have more complete chromite-ulvospinel solid solution with some grains having gradational zoning. Feldspathic gabbro (15388) has only ulvöspinel. Spinels in six pyroxene basalts show a large compositional gap, with chromites and ulvöspinels being highly enriched in their respective end members; ulvöspinel often mantles

[Reprinted from the Journal of the American Chemical Society, 96, 615 (1974).]
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner.

A74-31458

## Mechanisms of Unimolecular Gas-Phase γ-Hydrogen Rearrangements of Radical Cations. The Influence of Entropy as Opposed to Energy

*Sir:*

Those γ-hydrogen rearrangements commonly referred to as McLafferty-type rearrangements have probably attracted more attention than any other rearrangement encountered in mass spectrometry;[1] yet there is still considerable contention as to whether these ubiquitous rearrangements occur in a concerted or a stepwise manner.[2-6] We have sought to elucidate this persistent question by applying the powerful capabilities of field ionization kinetics (FIK)[7-9] to a study of two complementary γ-hydrogen rearrangements in hexanal.[10,11] The FIK technique[7-9] allows rates of unimolecular gas-phase reactions of (radical-) cations to be measured over a time range extending from $10^{-12}$ or $10^{-11}$ sec to $10^{-6}$ or $10^{-5}$ sec.

The rates of formation of *m/e* 45 ($C_2H_3DO$) and *m/e* 57

(1) (a) F. W. McLafferty, *Anal. Chem.*, **28**, 306 (1956); (b) *ibid.*, **31**, 82 (1959).
(2) F. P. Boer, T. W. Shannon, and F. W. McLafferty, *J. Amer. Chem. Soc.*, **90**, 7239 (1968).
(3) C. Fenselau, J. L. Young, S. Meyerson, W. R. Landis, E. Selke, and C. Leitch, *J. Amer. Chem. Soc.*, **91**, 6847 (1969).
(4) P. B. Fairweather and F. W. McLafferty, *Org. Mass. Spectrom.*, **2**, 755 (1969).
(5) P. Brown, A. H. Albert, and G. R. Pettit, *J. Amer. Chem. Soc.*, **92**, 3212 (1970).
(6) G. S. Smith and F. W. McLafferty, *Org. Mass Spectrom.*, **5**, 483 (1971).
(7) (a) H. D. Beckey, *Z. Naturforsch. A*, **16**, 505 (1961); *ibid.*, **21**, 1920 (1965); H. D. Beckey, H. Hey, K. Levsen, and G. Tenschert, *Int. J. Mass Spectrom. Ion Phys.*, **2**, 101 (1969).
(8) P. J. Derrick and A. J. B. Robertson, *Proc. Roy. Soc., Ser. A.*, **324**, 491 (1971).
(9) A. M. Falick, P. J. Derrick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, **12**, 101 (1973).
(10) R. J. Leidtke and C. Djerassi, *J. Amer. Chem. Soc.*, **91**, 6814 (1969).
(11) S. Meyerson, C. Fenselau, J. L. Young, W. R. Landis, E. Selke, and L. C. Leitch, *Org. Mass Spectrom.*, **3**, 689 (1970).

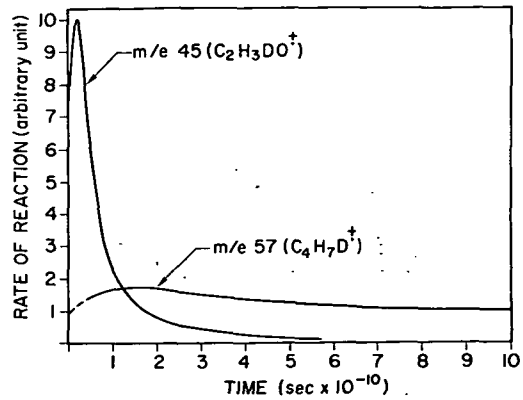



**Figure 1.** The dependence on time of the rates of formation of *m/e* 45 ($C_2H_3DO\cdot^+$) and *m/e* 57 ($C_4H_7D\cdot^+$) from hexanal-*4,4-d*$_2$ within the time range $10^{-11}$–$10^{-9}$ sec following FI. The rate of formation of *m/e* 57 ($C_4H_7D\cdot^+$) at the very shortest times can only be estimated due to the simultaneous formation of *m/e* 57 ($C_3H_5O^+$) by field dissociation.

($C_4H_7D$) from hexanal-*4,4-d*$_2$ following field ionization (FI) are shown in Figure 1 as functions of time over the range $10^{-11}$–$10^{-9}$ sec. The means by which these curves were obtained have been described.[9,12,13] The formation both of *m/e* 45 ($C_2H_3DO$) and of *m/e* 57 ($C_4H_7D$) involves $\gamma$-D rearrangement and $\beta$ cleavage. The hexanal-*4,4-d*$_2$ species was studied so as to distinguish between $\gamma$ transfer and transfer from other sites.[14] There is no H/D randomization prior to fragmentation at $10^{-11}$–$10^{-9}$ sec (ref 14). It is evident from Figure 1 that the relative rates of these two reactions are sensitive functions of time. At times of a few $\times 10^{-11}$ sec the rate of formation of *m/e* 45 ($C_2H_3DO$) exceeds the rate of formation of *m/e* 57 ($C_4H_7D$) by almost 1 order of magnitude. At times of the order of $10^{-10}$ sec, however, the formation of *m/e* 57 ($C_4H_7D$) is the faster reaction. The trend for the formation of *m/e* 57 ($C_4H_7D$) to be faster at longer times is supported by measurements[15] of the metastable transitions in the first field free region (corresponding to reaction at $\sim 10^{-6}$ sec after ionization). The transition $C_6H_{10}D_2O\cdot^+ \rightarrow C_4H_7D\cdot^+$ is *at least* 2 orders of magnitude more intense than $C_6H_{10}D_2O\cdot^+ \rightarrow C_2H_3DO\cdot^+$ (which cannot be detected at all).

We assume that the formation of *m/e* 45 ($C_2H_3DO$) and the formation of *m/e* 57 ($C_4H_7D$) are competing reactions of the ground electronic state. The average internal excitation energy available through FI is probably of the order of tenths of an eV.[16] The photoelectron spectrum of hexanal suggests that the ground electronic state may be an isolated state approximately 1.5 eV below the first excited state.[17] Reactions at the shortest times can be assumed to involve the most energetic reactant ions, and reactions at longer times the less energetic. Thus with the most energetic reactant ions, the rate of formation of *m/e* 45 ($C_2H_3DO$) exceeds the rate of formation of *m/e* 57 ($C_4H_7D$). It is at the lowest energies that energy considerations will be most influential in directing the competition between the reactions. The results suggest, therefore, that the formation of *m/e* 57 ($C_4H_7D$) is energetically more favorable than the formation of *m/e* 45 ($C_2H_3DO$). This is consistent with the observation in electron impact (EI) studies[10] that on lowering the electron energy the intensity of *m/e* 57 ($C_4H_7D$) increases relative to the intensity of *m/e* 45 ($C_2H_3DO$). The higher rate of formation of *m/e* 45 ($C_2H_3DO$) at the highest energies, however, is not readily explicable on energetic grounds. We suggest that the explanation lies in the realm of entropy. We suggest that the formation of *m/e* 45 ($C_2H_3DO$), although discriminated against by the energy factors, is favored by the entropy factors and that it is the entropy factors which direct the competition between the reactions at the shortest times ($\sim 10^{-11}$ sec).[18] In terms of transition-state theory, we are suggesting that the formation of *m/e* 57 ($C_4H_7D$) has the lower activation energy but that the formation of *m/e* 45 ($C_2H_3DO$) has the more favorable (*i.e.*, the less negative) activation entropy.

The kinetic results pose considerable mechanistic difficulties. For the present let us consider that both reactions are initiated by a common $\gamma$-D transfer step. This interpretation is consistent with recent trends.[5,6,11] Simple $\beta$ cleavage at times of the order of $10^{-11}$ sec could form the enol fragment *m/e* 45 ($C_2H_3DO$); at longer times $\beta$ cleavage *together with rearrangement* within the hydrocarbon chain could form *m/e* 57 ($C_4H_7D$) with the 2-butene structure. These mechanisms satisfy our energy and entropy criteria. The formation of *m/e* 57 ($C_4H_7D$) should be favored by energy considerations,[19] as the ionization potential (IP) of vinyl alcohol has been estimated as 9.5 eV[20] and that of 2-butene has been measured as 9.1 eV.[21] There is, however, an objection to these mechanisms. Simple $\beta$ cleavage subsequent to $\gamma$-hydrogen transfer at times of the order of $10^{-11}$ sec would be expected to form *both* *m/e* 45 ($C_2H_3DO$) and *m/e* 57 ($C_4H_7D$ with the 1-butene structure) at comparable rates. This is not what is observed (Figure 1). The IP of 1-butene has been measured as 9.6 eV.[21] The situation would be that of charge competition as discussed by Meyerson, *et al.*[11] The analogous reactions in heptanal induced by FI have the same kinetic characteristics as the reactions in hexanal.[22] Here the IP of the 1-pentene molecule competing for the charge is 9.5 eV.

Alternatively the natures of the cyclic transition states involved in the two initial $\gamma$-D transfers could differ. An attractive possibility is that the observed reactions correspond to stepwise and concerted mechanisms discussed so extensively in the literature.[2-6,11] Chemical intuition demands that the concerted mechanism

(12) J.-P. Pfeifer, A. M. Falick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, **11**, 345 (1973).
(13) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, **94**, 6794 (1972).
(14) (a) P. J. Derrick, A. M. Falick, and A. L. Burlingame, 20th Annual Conference Mass Spectrometry Allied Topics, Dallas, Texas, June 1972, Paper No. H4; (b) P. Brown and C. Fenselau, *Org. Mass Spectrom.*, 7, 305 (1973).
(15) See, for example, L. P. Hills and J. H. Futrell, *Org. Mass Spectrom.*, 5, 1019 (1971).
(16) G. Tenschert and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, 7, 97 (1971).
(17) A. W. Potts and W. C. Price, private communication.
(18) See P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, **95**, 437 (1973). For a fuller discussion, see P. J. Derrick and A. L. Burlingame, *Accounts Chem. Res.*, in press.
(19) D. P. Stevenson, *Discuss. Faraday Soc.*, No. 10, 35 (1951).
(20) S. Meyerson and J. D. McCollum, *Advan. Anal. Chem. Instrum.*, 2, 179 (1963).
(21) J. L. Franklin, J. G. Dillard, H. M. Rosenstock, J. T. Herron, K. Draxl, and F. H. Field, "Ionization Potentials, Appearance Potentials, and Heats of Formation of Gaseous Positive Ions," NSRDS-NBS 26, U. S. Government Printing Office, Washington, D. C., 1969.
(22) P. J. Derrick, A. M. Falick, and A. L. Burlingame, unpublished data.

represent the faster process leading to *m/e* 45 ($C_2H_3DO$). The slower process *via* an intermediate would lead to *m/e* 57 ($C_4H_7D$) presumably with the 2-butene structure. Calculations by Boer, Shannon, and McLafferty[2] indicate that the transition state in the stepwise mechanism is energetically preferable to that in a concerted mechanism. These mechanisms would demand that the transition state for the concerted mechanism be "looser" than that in the stepwise.[23] This would probably mean that the transition state in the concerted mechanism does not need to be exactly planar. The weakly bonding and extended $\beta$ bond in the concerted transition state might result in the internal rotations of the reactant being less constrained than they would be in the stepwise transition state. This could give the concerted transition state a higher entropy.[24]

The kinetic results leave little doubt that there are two distinct processes in hexanal involving $\gamma$-hydrogen transfer and $\beta$ cleavage. It does not seem useful to refer to two mechanistically distinct reactions by the same name. We would suggest that it is time to give a more precise and therefore more meaningful definition to the term "McLafferty rearrangement."

**Acknowledgment.** Financial support for this work was provided by the National Aeronautics and Space Administration (Grant NGL 05-003-003) and the National Science Foundation (Grant NSF GP-38389X).

(23) (a) G. M. Wieder and R. A. Marcus, *J. Chem. Phys.*, 37, 1835 (1962); (b) B. S. Rabinovitch and D. W. Setser, *Advan. Photochem.*, 3, 1 (1964).

(24) H. E. O'Neal and S. W. Benson, *J. Phys. Chem.*, 71, 2903 (1967).

(25) Address correspondence to this author at the Department of Chemistry, University College, University of London, London WC1H OAJ, England.

**P. J. Derrick,*[25] A. M. Falick, A. L. Burlingame**
*Space Sciences Laboratory, University of California*
*Berkeley, California 94720*
*Received July 5, 1973*

[Reprinted from the Journal of the American Chemical Society, 96, 4909 (1974).]
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner.

A75-20514

# Progressive Specific Hydrogen Rearrangements Quenched by Molecular Ion Fragmentation in 2-Methylpropene. Mass Spectral Serendipity[1]

**P. J. Derrick*[2a] and A. L. Burlingame[2b]**

*Contribution from the Space Sciences Laboratory, University of California, Berkeley, California 94720.* *Received September 7, 1972*

**Abstract:** The loss of a methyl radical from 2-methylpropene, 2-methylpropene-*1,1*-$d_2$, and 2-methyl-$d_3$-propene-*3,3,3*-$d_3$ has been studied at times from $10^{-11}$ to $10^{-6}$ sec following field ionization (FI). Considerable rearrangement of the H and D in the deuterated species occurs prior to fragmentation even at times as short as a few $\times 10^{-11}$ sec. The rearrangement leads to almost complete H–D randomization prior to fragmentation at $7 \times 10^{-10}$ sec. The observed H–D randomization may be rationalized in terms of successive isomerizations *via* 1,3 allylic hydrogen shifts. The electron impact (EI) mass spectra of 2-methylpropene, 2-methylpropene-*1,1*-$d_2$, and 2-methyl-$d_3$-propene-*3,3,3*-$d_3$ have been measured at high and low ionizing energies and are discussed in relation to the FI results.

It is recognized that the unimolecular reaction systems set up by gas-phase ionization of alkenes are typically extremely complex.[3] The mass spectra of isotopically labeled species show that extensive molecular rearrangement leading to randomization of the labels tends to occur either prior to or as an integral part of fragmentation.[4–7] The nature of the rearrangement processes remains, however, largely a matter of conjecture, even for the smallest molecules such as propenes and butenes. It has been proposed that the processes leading to hydrogen randomization in the propene ion involve only 1,3 shifts.[6b,c] Other workers have suggested that both 1,2 and 1,3 shifts are operative.[6d] Yet again it has been suggested that randomization in propene, butene, and pentene radical-ions is predominantly due to 1,2-hydrogen shifts.[6e] 1,3-Allylic hydrogen shifts have been envisaged to occur in the $\Delta^{4(8)}$-menthene[7] and methylcyclohexene[8] radical-ions. McLafferty, *et al.*,[9] have proposed that closely analogous 1,3-hydrogen shifts occur in the enolic $C_3H_6O\cdot^+$ ion. Conversely, however, it has been suggested[10] that whereas 1,2-, 1,4-, and 1,5-hydrogen shifts are in general facile unimolecular reactions of radical-cations, 1,3-

(1) The work was performed in collaboration with Dr. A. M. Falick who is also responsible for the computation of lifetimes. See J.-P. Pfeifer, A. M. Falick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, **11**, 345 (1973).

(2) (a) Present address: Department of Chemistry, University College, University of London, London WC1H OAJ, England; (b) John Simon Guggenheim Memorial Fellow, 1970–1972.

(3) See H. Budzikiewicz, C. Djerassi, and D. H. Williams, "Mass Spectrometry of Organic Compounds," Holden-Day, San Francisco, Calif., 1967.

(4) W. A. Bryce and P. Kebarle, *Can. J. Chem.*, **34**, 1249 (1956).

(5) J. H. Beynon, "Mass Spectrometry and Its Application to Organic Compounds," Elsevier, Amsterdam, 1960, pp 262–263.

(6) (a) G. K. Helmkamp and N. Schnautz, *J. Org. Chem.*, **24**, 529 (1959); (b) H. H. Voge, C. D. Wagner, and D. P. Stevenson, *J. Catal.*, **2**, 58 (1963); (c) W. H. McFadden, *J. Phys. Chem.*, **67**, 1074 (1963); (d) S. R. Smith, R. Schor, and W. P. Norris, *ibid.*, **69**, 1615 (1965); (e) B. J. Millard and D. F. Shaw, *J. Chem. Soc. B*, 664 (1966).

(7) D. S. Weinberg and C. Djerassi, *J. Org. Chem.*, **31**, 115 (1966).

(8) T. H. Kinstle and R. E. Stark, *J. Org. Chem.*, **32**, 1318 (1967).

(9) F. W. McLafferty, D. J. McAdoo, J. S. Smith, and R. Kornfield, *J. Amer. Chem. Soc.*, **93**, 3720 (1971).

(10) M. T. Bowers, D. H. Aue, and D. D. Elleman, *J. Amer. Chem. Soc.*, **94**, 4255 (1972).

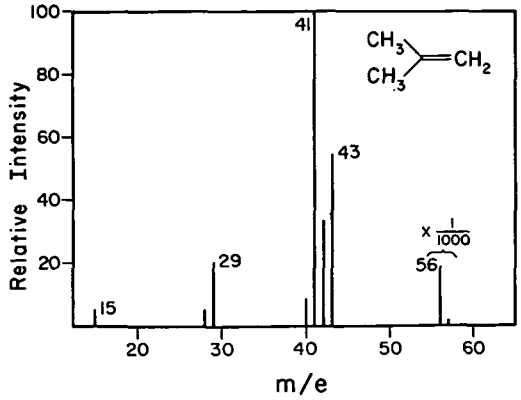



**Figure 1.** The "normal" FI mass spectrum of 2-methylpropene.

hydrogen shifts are not.[11,12] The earlier workers tended to neglect the possibility of carbon skeletal rearrangements. More recent work,[13-15] however, supports an early suggestion[5] that at least some alkene radical-ions can undergo skeletal rearrangement. The 1-butene radical-ion, for example, rearranges to the 2-methylpropene carbon skeleton structure even at very low energies[13,14] (photoionization with the argon resonance lines 11.6 and 11.8 eV). Carbon skeletal rearrangements of alkene radical-ions are not at all understood. Taking an overall view, the present understanding of the unimolecular gas-phase rearrangement and fragmentation of alkene radical-ions could justifiably be described as confused. It has become apparent that insight into the nature of processes occurring at rates too rapid to be directly observed by electron impact (EI) mass spectrometry may be gained by proper application of the field ionization kinetics (FIK) technique.[16,17] An earlier FIK study[18] provided considerable insight into the nature of the rearrangements and fragmentations of the cyclohexene radical-cation. Observed hydrogen randomization could be explained in terms of rapid 1,3 allylic hydrogen shifts around an intact cyclohexene ring. We now report FIK measurements on 2-methylpropene, 2-methylpropene-*1,1-d*$_2$ (**1**), and 2-methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**). There is good evidence[14,19] that the 2-methylpropene radical-ion does not undergo skeletal rearrangement at low energies, so that any observed hydrogen randomization can be discussed with some confidence in terms of hydrogen shifts of some kind or another.

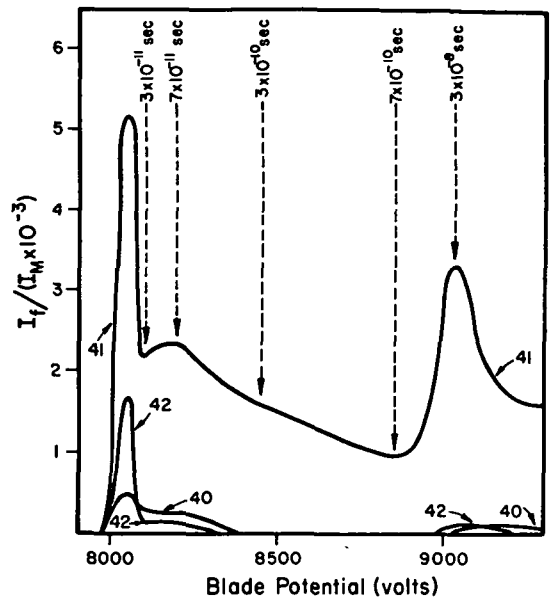



**Figure 2.** Measured curves of ion currents $I_f$ as a function of blade potential for *m/e* 41 $(M - CH_3)^+$, *m/e* 40 $(M - CH_4)\cdot^+$, and *m/e* 42 $([M + 1] - CH_3)^+$ following FI of 2-methylpropene. $I_M$ is the ion current of the molecular ion $M^+$ in the "normal" FI mass spectrum. Experimental data points are omitted due to their large number and low degree of scatter. The times quoted refer to the formation of *m/e* 41 $(M - CH_3)^+$.

## Experimental Section

The FIK measurements were made on a modified Du Pont (C.E.C.) 21-110B double focusing mass spectrometer.[20] The standard Du Pont combination FI/EI combination ion source incorporating a blade emitter was employed. The electron collimating magnet was removed to avoid mass discrimination. Source temperature was maintained at 200° during all of the reported experiments.

The theory[20-26] behind the field ionization kinetic measurements described here and a detailed description of the method employed[20] in obtaining our results have been presented previously, so that only a few brief comments need be given here. The magnetic analyzer was focused on the fragment ion of interest and the blade potential $V_B$ was scanned from approximately 8 to 10 kV in 10-V steps. The resulting quasicontinuous curves could be converted to curves of fragment ion current (or rate of fragment ion formation) *vs.* time.[1,20] The shapes of these curves are dependent on source temperature[27] and conditions of the blade surface.[21,28]

EI mass spectra were measured on a G.E.C.-A.E.I. MS 902 mass

(11) Direct quantitative data on hydrogen shifts in carbonium ions in the liquid phase have recently been provided (M. Saunders and J. J. Stofko, *J. Amer. Chem. Soc.*, **95**, 252 (1973)). Rearrangement of the 2,4-dimethyl-2-pentyl cation does *not* proceed *via* successive 1,2-hydrogen shifts. Rather, 1,3-hydrogen shift occurs. The much studied propyl cation rearranges by 1,3 instead of by successive 1,2-hydrogen shifts. For a review, see J. L. Fry and G. J. Karabatsos in "Carbonium Ions," Vol. II, G. A. Olah and P. v. R. Schleyer, Ed., Wiley-Interscience, New York, N. Y., 1970, pp 521–572.

(12) 1,3-Hydrogen shifts are well known in free-radical chemistry. In contrast, the efforts to securely establish the occurrence of a 1,2-hydrogen shift within a free-radical have as yet been unsuccessful. For a review see W. A. Pryor, "Free Radicals," McGraw-Hill, New York, N. Y., 1966.

(13) G. G. Meisels, J. Y. Park, and B. G. Giessner, *J. Amer. Chem. Soc.*, **91**, 1555 (1969).

(14) L. W. Sieck, S. G. Lias, L. Hellner, and P. Ausloos, *J. Res. Nat. Bur. Stand., Sect. A*, **76**, 115 (1972).

(15) P. S. Gill, Y. Inel, and G. G. Meisels, *J. Chem. Phys.*, **54**, 2811 (1971).

(16) We define data as to the kinetics of unimolecular gas-phase reactions at times from $10^{-11}$ sec (or in some cases $10^{-12}$ sec) to $10^{-5}$ sec following field ionization as "field ionization kinetics." See P. J. Derrick, A. M. Falick, and A. L. Burlingame, 20th Annual Conference on Mass Spectrometry and Allied Topics, Dallas, Texas, June 1972, Paper No. H4.

(17) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *Advan. Mass Spectrom.*, **6**, 877 (1974).

(18) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, **94**, 6794 (1972).

(19) S. G. Lias and P. Ausloos, *J. Res. Nat. Bur. Stand., Sect. A*, **75**, 59 (1971).

(20) A. M. Falick, P. J. Derrick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, **12**, 101 (1973).

(21) P. J. Derrick and A. J. B. Robertson, *Proc. Roy. Soc., Ser. A*, **324**, 491 (1971).

(22) H. D. Beckey, *Z. Naturforsch. A*, **16**, 505 (1961).

(23) H. D. Beckey and H. Knoppel, *Z. Naturforsch. A*, **21**, 1920 (1966).

(24) H. D. Beckey, H. Hey, K. Levsen, and G. Tenschert, *Int. J. Mass Spectrom. Ion Phys.*, **2**, 101 (1969).

(25) H. G. Metzinger and H. D. Beckey, *Z. Phys. Chem. (Frankfurt am Main)*, **52**, 27 (1967).

(26) P. J. Derrick, Ph.D. Thesis, University of London, 1969.

(27) K. Levsen and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, **9**, 51 (1972).

(28) P. J. Derrick and A. J. B. Robertson, *Int. J. Mass. Spectrom. Ion Phys.*, **10**, 315 (1973).

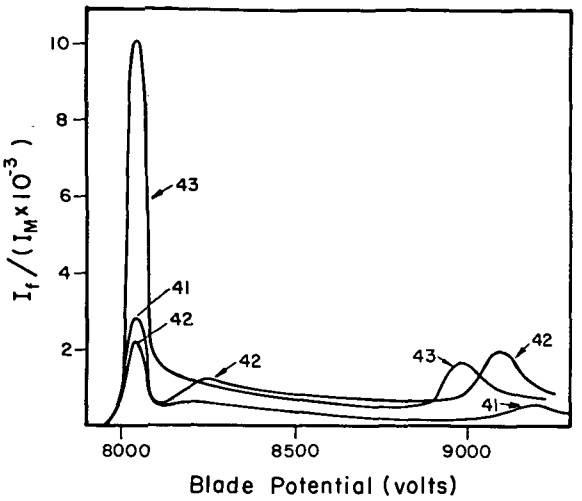



**Figure 3.** Measured curves of ion currents $I_f$ as a function of blade potential for *m/e* 41 $(M - CD_2H)^+$, *m/e* 42 $(M - CDH_2)^+$, and *m/e* 43 $(M - CH_3)^+$ following FI of 2-methylpropene-*1,1-d*$_2$ (**1**). $I_M$ is the ion current of the molecular ion $M^+$ in the "normal" FI mass spectrum. Experimental data points are omitted due to their large number and low degree of scatter.

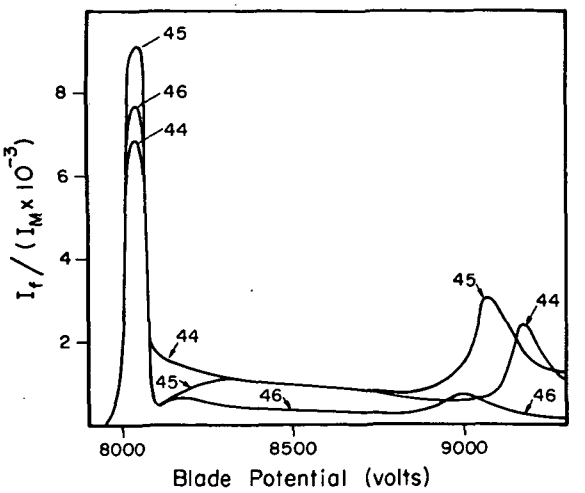



**Figure 4.** Measured curves of ion currents $I_f$ as a function of blade potential for *m/e* 44 $(M - CD_3)^+$, *m/e* 45 $(M - CD_2H)^+$, and *m/e* 46 $(M - CDH_2)^+$ following FI of 2-methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**). $I_M$ is the ion current of the molecular ion $M^+$ in the "normal" FI mass spectrum. Experimental data points are omitted due to their large number and low degree of scatter.

spectrometer on-line to an XDS Sigma 7 computer.[29] Source temperature was maintained at 50°.

Both deuterated propenes **1** and **2** supplied by Merck, Sharp and Dohme Ltd., Montreal, were found by mass spectrometry to have isotopic purities of 98 atom %.

## Results

The "normal" FI mass spectrum of 2-methylpropene is given in Figure 1. The "normal" FI mass spectrum obtained with a double focusing analyzer displays only the ionic products of reactions occurring on the blade surface or in the immediate vicinity of the blade.[20] The "normal" spectrum is obtained when the blade voltage is approximately equal to the normal accelerating voltage.[30]

Figure 2 gives the ion currents for the fragment ions *m/e* 40, 41, and 42 as a function of blade potential. The ion currents at blade voltages greater than about 8100 V are due entirely to gas-phase unimolecular decompositions.[20] The ion currents at $V_B < 8100$ V represent ions formed on and very close to the blade surface. The maxima in the ion currents in Figure 3 at $V_B > 9000$ V represent fragmentation occurring between the focus electrodes.[12,20] The ions *m/e* 40, 41, and 42 are the only fragments formed between the blade and the cathode outside of the high field (*i.e.*, at times $>\sim 3 \times 10^{-11}$ sec) following FI of 2-methylpropene.

Following field ionization of the deuterated species 2-methylpropene-*1,1-d*$_2$ (**1**), three fragments *m/e* 41, 42, and 43 are formed at significant intensities between the blade and the cathode outside of the high-field region. Similarly 2-methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**) yields three fragments *m/e* 44, 45, and 46. The ion currents $I_f$ of these fragments as functions of blade potential $V_B$ are shown in Figures 3 and 4. The maxima at around 9000 V in both figures are as before due to fragmentation between the focus electrodes. The curves of ion current $I_f$ against blade potential $V_B$ between $V_B = 8100$ V and about $V_B = 8900$ V have been transformed to curves of ion current $I_f(t)$ against molecular ion lifetime in Figures 5 and 6. The ion current $I_f$ at $V_B < 8100$ V is not considered since these ions may arise from surface or high-field processes. The ratios of the currents $I_f(t)$ at any particular time within either Figure 5 or 6 are equal to a good approximation to the ratios of the rates of formation of the fragment ions.[20]

The electron impact (EI) mass spectra of 2-methylpropene, **1**, and **2** at low and high electron energies are presented in Figure 7.

The relative intensities of the fragments *m/e* 41, 42, and 43 from **1** and of *m/e* 44, 45, and 46 from **2** at different times after FI and the relative intensities of these ions in the low and high energy EI spectra are displayed in Table I. The FI intensities at times less than and including $3 \times 10^{-8}$ sec are based on Figures 3–6. The intensities at $10^{-6}$ sec following FI are based on metastable transitions in the field-free region between the ion source and the electric sector measured by the now well-established "metastable defocusing technique" of lowering the electric sector potential.[31]

## Discussion

**Field Ionization Kinetics.** Our interests in this FIK investigation of 2-methylpropene concern unimolecular reactions taking place in the gas phase beyond the influence of the very high external electric fields in the immediate vicinity of the blade.[32] Conclusions reached in the FI study can then be *tentatively* extrapolated to low energy (10–12 eV), and possibly high energy (70 eV),

(29) A. L. Burlingame in "Recent Developments in Mass Spectroscopy," K. Ogata and T. Hayakawa, Ed., University Park Press, Baltimore, Md., 1970, p 104.

(30) The blade voltage differs from the normal accelerating voltage by a small correcting term. See A. J. B. Robertson and B. W. Viney, *J. Chem. Soc. A*, 1843 (1966).

(31) (a) M. Barber and R. M. Elliott, 12th Annual Conference on Mass Spectrometry and Allied Topics, Montreal, June 1964, Paper No. 22; (b) L. P. Hills and J. H. Futrell, *Org. Mass Spectrom.*, 5, 1019 (1971).

(32) Generally speaking "normal" FI mass spectra are not reliable bases for mechanistic inference for *unimolecular gas-phase* reactions. The reactions contributing to the "normal" FI mass spectrum may be catalyzed by the emitter surface, often occur in a liquid phase at the surface, and generally are influenced by the very high field in the vicinity of the emitter. See, for example, P. Brown and C. Fenselau, *Org. Mass Spectrom.*, 7, 305 (1973); E. M. Chait and F. G. Kitson, *ibid.*, 3, 533 (1970).

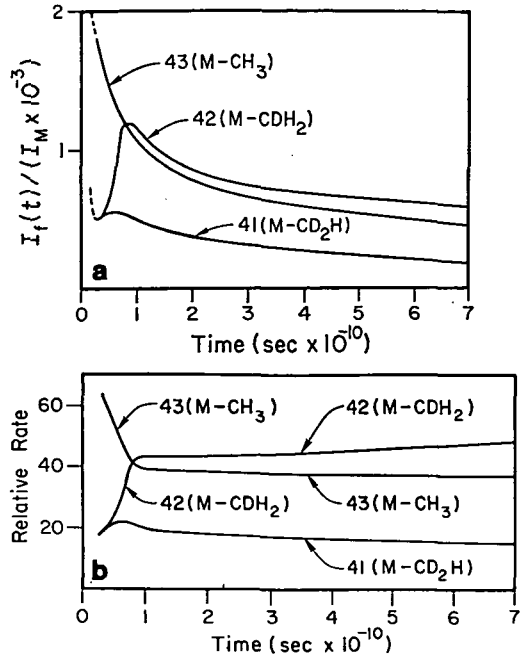



**Figure 5.** (a) Ion currents $I_f(t)$ due to *m/e* 41 $(M - CD_2H)^+$, *m/e* 42 $(M - CDH_2)^+$, and *m/e* 43 $(M - CH_3)^+$ from 2-methylpropene-*1,1-d*$_2$ (**1**) as a function of molecular ion lifetime following FI. The lifetimes are calculated from the blade potentials in Figure 3. $I_M$ is the ion current of the molecular ion $M^+$ in the "normal" FI mass spectrum. (b) Relative rates of formation of *m/e* 41 $(M - CD_2H)^+$, *m/e* 42 $(M - CDH_2)^+$, and *m/e* 43 $(M - CH_3)^+$ from 2-methylpropene-*1,1-d*$_2$ (**1**) as a function of molecular ion lifetime.

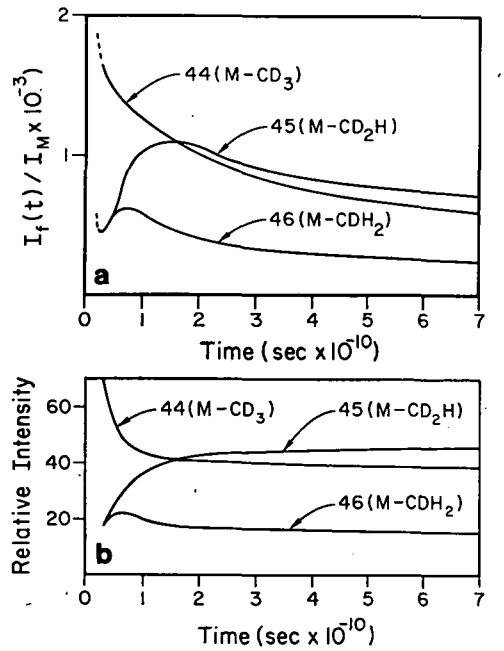



**Figure 6.** (a) Ion currents $I_f(t)$ due to *m/e* 44 $(M - CD_3)^+$, *m/e* 45 $(M - CD_2H)^+$, and *m/e* 46 $(M - CDH_2)^+$ from 2-methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**) as a function of molecular ion lifetime following FI. The lifetimes are calculated from the blade potentials in Figure 4. $I_M$ is the ion current of the molecular ion $M^+$ in the "normal" FI mass spectrum. (b) Relative rates of formation of *m/e* 44 $(M - CD_3)^+$, *m/e* 45 $(M - CD_2H)^+$, and *m/e* 46 $(M - CDH_2)^+$ from 2-methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**) as a function of molecular ion lifetime.

EI reaction systems. Following FI the parent ions move rapidly away from the blade surface in the high-potential gradient, and arrive at regions where the external

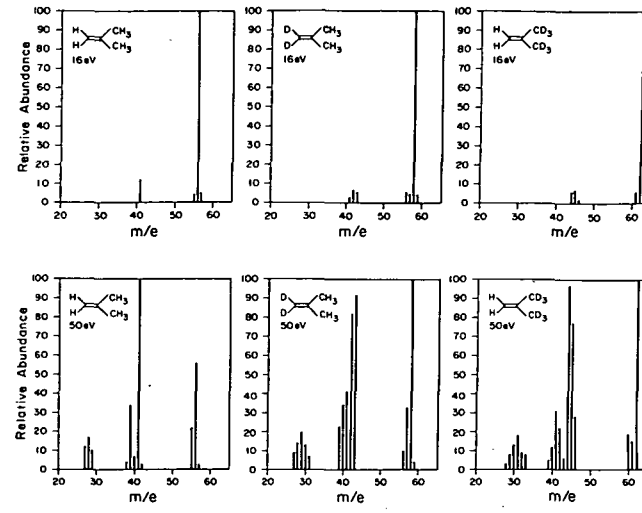



**Figure 7.** EI mass spectra of 2-methylpropene, 2-methylpropene-*1,1-d*$_2$ (**1**), and 2-methyl *d*$_3$-propene-*3,3,3-d*$_3$ (**2**) at 50 eV and nominally 16 eV ionizing electron energies. Ions with intensities of <1% relative to the most intense ion in a spectrum have been omitted.

electric fields are insignificant compared to the internal electric fields of the molecule within 2 or $3 \times 10^{-11}$ sec.[20,21] We are interested in fragment ions formed at times $> 3 \times 10^{-11}$ sec (Figures 2, 5, and 6). The formation of the fragment *m/e* 41 from 2-methylpropene at times $>3 \times 10^{-11}$ sec following FI is interpreted as the elimination of a methyl radical from the molecular ion. The less important (for the purposes of our discussion) fragments *m/e* 40 and 42 at times $>3 \times 10^{-11}$ sec are attributed to loss of methane from the molecular ion and loss of a methyl radical from the protonated species $(M + 1)^+$, respectively. Transitions $56^+ \rightarrow 40^+$ and $57^+ \rightarrow 42^+$ are observed at low intensities in the first field-free region. The formation of protonated molecular ions $(M + 1)^+$ in reactions at the surface is very common following FI.[33] Phenomenological rate constants $\bar{k}(t)$ have been calculated for the formation of *m/e* 41 $(M - CH_3)$ from 2-methylpropene (see Figure 8) using the relationship[20,33]

$$\bar{k}(t) = I_f(t)/I_M(t)\Delta t \qquad (1)$$

In an earlier paper,[18] we referred to $\bar{k}(t)$ as an "average rate constant." We now prefer the term "phenomenological rate constant," since this emphasizes that $\bar{k}(t)$ is basically an experimental quantity and, further, emphasizes the close analogy with phenomenological rate constants of ion-molecule reactions. Accordingly we refer to the theoretical rate constant $k$ describing the ideal situation in which all reactant molecules possess the same discrete amount of internal energy as a "microscopic rate constant." The quantities on the right-hand side of the relationship (eq 1) are as follows: $I_f(t)$ is the *m/e* 41 $(M - CH_3)$ ion current at time $t$, $I_M(t)$ is the ion current of undecomposed molecular ions *m/e* 58 (M) at that time, and $\Delta t$ is the time interval over which the fragment ions contributing to $I_f(t)$ are formed. The values of $I_f(t)$ at $10^{-11}$–$10^{-9}$ sec are taken from Figure 2. The value at $10^{-6}$ sec is the metastable intensity in the first field-free region. $I_M(t)$ is replaced by the molecular ion current $I_M$ in the "normal" FI mass

(33) See H. D. Beckey, "Field Ionization Mass Spectrometry," Pergamon Press, New York, N. Y., 1971.

Table I. Measured Relative Intensities of Fragment Ions Formed by Loss of a Methyl Radical from 2-Methylpropene-*1,1-d*$_2$ (1) and 2-Methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (2) at Various Times Following FI and Following EI. The Calculated Relative Intensities Assuming Complete Randomization of H–D Prior to Loss of Methyl Are Given for Comparison

| | 2-Methylpropene-*1,1-d*$_2$ (1) | | | 2-Methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (2) | | |
|---|---|---|---|---|---|---|
| | *m/e* | | | *m/e* | | |
| | 41 $(M - CD_2H)^+$ | 42 $(M - CDH_2)^+$ | 43 $(M - CH_3)^+$ | 46 $(M - CDH_2)^+$ | 45 $(M - CD_2H)^+$ | 44 $(M - CD_3)^+$ |
| Calculated (complete H–D randomization)[a] | 11 | 54 | 36 | 11 | 54 | 36 |
| Field ionization | | | | | | |
| $3 \times 10^{-11}$ sec | 18 | 18 | 63 | 18 | 18 | 65 |
| $1 \times 10^{-10}$ sec | 18 | 43 | 39 | 20 | 36 | 44 |
| $7 \times 10^{-10}$ sec | 15 | 48 | 37 | 15 | 46 | 39 |
| $3 \times 10^{-8}$ sec (at focus electrodes[b]) | 14 | 47 | 39 | 12 | 49 | 39 |
| $1 \times 10^{-6}$ sec (metastables in first field-free region) | 10 | 51 | 39 | 11 | 55 | 34 |
| Electron impact | | | | | | |
| 50 eV[c] | 19 | 38 | 43 | 14 | 38 | 48 |
| Nominally 16 eV | 13 | 49 | 38 | 12 | 50 | 38 |

[a] Statistical mixture of all possible 2-methylpropene isotopic isomers. [b] Small errors will be present in these figures due to contributions from fragments formed between the blade and the cathode (*i.e.*, at times $\sim 10^{-9}$ sec). [c] With 2-methylpropene-*1,1-d*$_2$ (1), the measured intensity of *m/e* 41 probably contains a significant contribution from $C_3D_2H^+$ $(M - CH_3 - H_2)^+$.

spectrum. The estimation of $\Delta t$ has been described.[1,12,20] A linear dependence of the phenomenological rate constants on time such as we find for the formation of *m/e* 41 $(M - CH_3)$ has been found for many other unimolecular gas-phase reactions induced by field ionization.[12,24,33,34] As far as can be assessed, the linearity in Figure 8 extends from $10^{-10}$ to $10^{-6}$ sec. At times less than $1 \times 10^{-10}$ sec, $\bar{k}(t)$ is approximately independent of time. The variation in the magnitude of the phenomenological rate constant $\bar{k}(t)$ from $10^7$ to $10^3$ $sec^{-1}$ would suggest that a *distribution* of microscopic rate constants $k$ is necessary for an adequate description of the reaction. A distribution of microscopic rate constants $k$ might arise if different amounts of internal energy $E$ were imparted to different molecules by the ionization process. More energetic ions react at higher rates and therefore at shorter times. As time progresses, so the reaction observed is that of progressively less energetic ions. Consequently the observed reaction rate and the phenomenological rate constant decline with time. The time independence of $\bar{k}(t)$ at the shortest times can be explained as follows. The explanation derives from Beckey and colleagues.[34c,d,35] Assuming the basic tenets of the quasi-equilibrium theory[36] to be valid and making certain reasonable assumptions[34c,d] as to the form of the distribution of microscopic rate constants $k$, it can be shown[34c,d] that the phenomenological rate constant $\bar{k}(t)$ will be approximately independent of time at times less than the reciprocal of the maximum microscopic rate constant $k_{max}$ in the distribution of microscopic rate constants. This happens since at these times reaction can be described by just the one maximum microscopic rate constant $k_{max}$. According to this theory, the change in slope at about $1 \times 10^{-10}$ sec in Figure 8

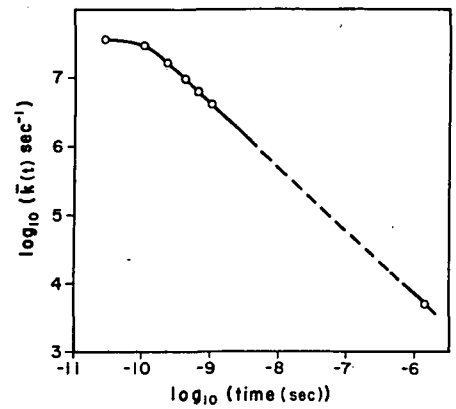



Figure 8. The logarithm of the phenomenological rate constant $\bar{k}(t)$ for the formation of *m/e* 41 $(M - CH_3)^+$ from 2-methylpropene as a function of the logarithm of the molecular ion lifetime following FI. Correction has been made for the increase in total emission on raising the blade potential.

must represent the reciprocal of $k_{max}$, so that $k_{max}$ for the loss of a methyl radical from the 2-methylpropene molecular ion (outside of the very high field region) is about $5 \times 10^9$ $sec^{-1}$ (a symmetry factor of 2 must be considered). This figure represents the maximum possible rate of reaction given the amount of excitation energy available from FI under the particular conditions employed. The reaction almost certainly proceeds more rapidly at higher energies (*e.g.*, following 70 eV EI).

We attribute the fragments formed from the deuterated species at times from $3 \times 10^{-11}$ to $10^{-9}$ sec to loss of methyl radicals from the molecular ions (and isotopic isomers thereof). Contributions from other fragmentation processes can be safely neglected. The fragments are thus identified as $(M - CD_2H)$ *m/e* 41, $(M - CDH_2)$ *m/e* 42, and $(M - CH_3)$ *m/e* 43 from **1** (Figures 4 and 6), and $(M - CD_3)$ *m/e* 44, $(M - CD_2H)$ *m/e* 45, and $(M - CDH_2)$ *m/e* 46 from **2** (Figures 4 and 6). The observation of three fragment ions with each of the deuterated species **1** and **2** corresponding to a single fragment $(M - CH_3)^+$ with the undeuterated species establishes that considerable rearrangement occurs

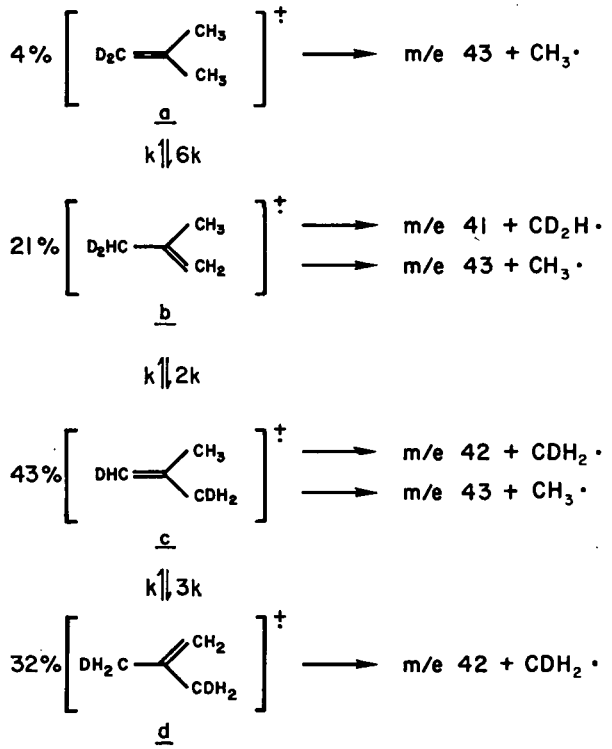
(34) (a) G. Tenschert and H. D. Beckey, *Ber. Bunsen. Phys. Chem.*, 72, 1039 (1968); (b) H. D. Beckey, *Z. Naturforsch. A*, 26, 1243 (1971); (c) G. Tenschert and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, 7, 97 (1971); (d) K. Levsen and H. D. Beckey, *ibid.*, 7, 341 (1971); (e) *ibid.*, 9, 51 (1972); (f) *ibid.*, 9, 63 (1972); (g) P. J Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, 95, 437 (1973); (h) P. J. Derrick, A. M. Falick, S. Lewis, and A. L. Burlingame, *Org. Mass Spectrom.*, 7, 887 (1973).

(35) H. D. Beckey, private communication.

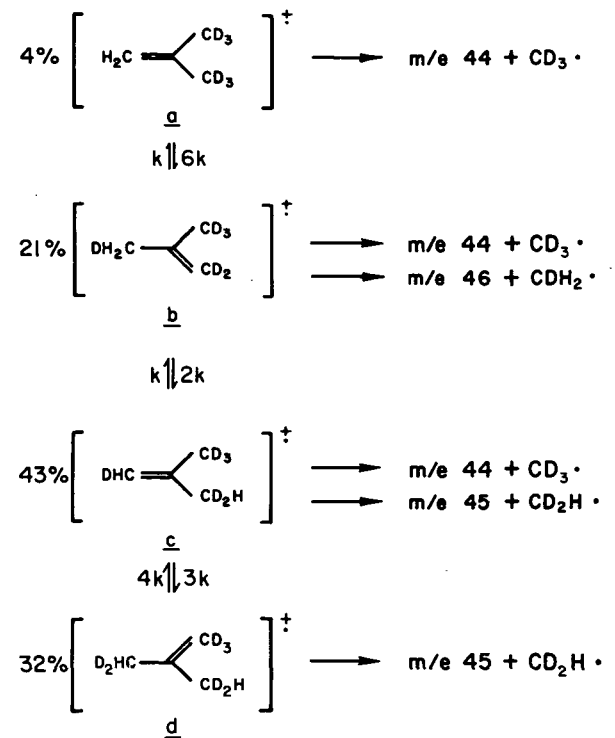
(36) H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig, and H. Eyring, *Proc. Nat. Acad. Sci. U. S.*, 38, 667 (1952).

prior to fragmentation. It is evident (Figures 5 and 6 and Table I) that the relative intensities of these fragment ions from **1** and **2** are sensitive functions of time. We suggest that the progress of hydrogen rearrangement prior to fragmentation can be monitored by observation of the manner in which the different fragmentations depend upon time. Consideration of Figures 5 and 6 and Table I leads to the conclusion that rearrangement and fragmentation occur in the manner depicted in Schemes I and II. Rearrangement is proposed to occur *via* 1,3-hydrogen shifts. The rate constants for the various isomeric interconversions are given in Schemes I and II in terms of the rate constant for a 1,3 hydrogen or deuterium shift. We consider the rates of hydrogen and deuterium transfer to be equal since the results suggest that any isotope effect is small.[37] The relative amounts of the different isomers in the statistical mixtures of the isomers are given as percentages in Schemes I and II. The mechanistic Schemes I and II satisfactorily explain the kinetic results. The discussion is similar with both **1** and **2**, so we shall consider only **1**. The relative intensity of *m/e* 43 (M − $CH_3$) declines with increasing time from a level considerably in excess of the statistical value at $3 \times 10^{-11}$ sec to approximately the statistical value at $1 \times 10^{-6}$ sec (Figure 5b and Table I). By statistical value we mean the relative intensity expected given a statistical mixture of all possible isotopic isomers a, b, c, and d and an absence of isotope effects. We define such a statistical mixture of all possible isotopic isomers of the structural species as H–D randomization. We attribute the declining intensity of *m/e* 43 (M − $CH_3$) with time to the progressive rearrangement of the molecular ion a to the isomers b, c, and d. Whereas all fragmentation of a gives *m/e* 43 (M − $CH_3$), only half the fragmentation of b and c leads to *m/e* 43 (M − $CH_3$). The fragmentation of d does not give *m/e* 43 (M − $CH_3$) at all. That the relative intensity of *m/e* 43 (M − $CH_3$) at the shortest accessible time ($3 \times 10^{-11}$ sec) is 63%, rather than 100% as would be the case if only the molecular ion a were present, indicates that a undergoes considerable rearrangement even within the short time interval $3 \times 10^{-11}$ sec.

**Scheme I.** 2-Methylpropene-*1,1-d*$_2$




**Scheme II.** 2-Methyl-*d*$_3$-propene-*3,3,3-d*$_3$ (**2**)



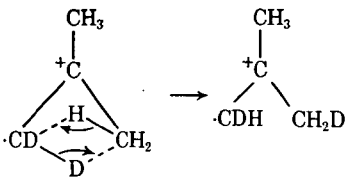

Consider now the *m/e* 41 (M − $CD_2H$) fragment. The relative intensity rises from 18% at $3 \times 10^{-11}$ sec to reach a maximum of 21% at $6 \times 10^{-11}$ sec. The maximum is obvious in Figure 5. Thereafter the relative intensity declines toward the statistical value of 11%. The fragment *m/e* 41 (M − $CD_2H$) is formed only from the isomer b according to our mechanistic scheme. A maximum in the relative intensity of *m/e* 41 (M − $CD_2H$) implies a maximum in the proportion of b in the isotopic mixture. The initial field ionization forms the molecular ion a. The rate constant $6k$ for the rearrangement of a to b is three times greater than the rate constant $2k$ for the conversion of b to c and six times greater than the rate constant $k$ for the reverse reaction b → a. At short times when the intensity of a is high and that of b is low (at very short times we can neglect c), the rate of formation of b will considerably exceed its rate of destruction. It must be expected that the relative intensity of b will rise rapidly and exceed the statistical value. There will come a time when the rate of formation of b, which depends on the intensity of a, will no longer exceed the rate of destruction of b, which depends on the intensity of b itself. The relative intensity of b will be at a maximum at that time[38] and

(37) It is evident (Table I) that the progress of the relative intensities of the fragment ions toward the statistical figures is very similar for the two deuterated species. If there were a significant isotope effect favoring hydrogen transfer over deuterium transfer, we would expect **1** to proceed to randomization more rapidly than **2** (see Schemes I and II). We conclude therefore that any kinetic isotope effect is small. Lack of kinetic isotope effects is consistent with the proposed mechanism which can be seen as having a highly symmetric transition state.

(38) It is possible to analyze the situation at the time $6 \times 10^{-11}$ sec when the maximum occurs. We neglect isotope effects on the frag-

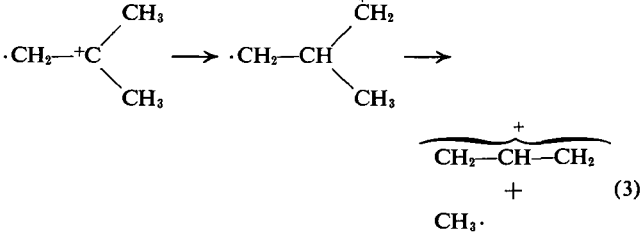
decline thereafter to the statistical value. Consider now *m/e* 42 (M − $CDH_2$). At no time of the order of $10^{-11}$ or $10^{-10}$ sec does the relative intensity exceed the statistical value 54%. The relative intensity rises steadily from 18% at $3 \times 10^{-11}$ sec to 48% at $7 \times 10^{-10}$ sec. That the relative intensity of *m/e* 42 (M − $CDH_2$) remains below the statistical value at all times of the order of $10^{-11}$ and $10^{-10}$ sec is readily rationalized on the basis of Scheme I. The fragment is formed from the isomer d and to a lesser extent from the isomer c. Neither c nor d can be formed directly from a, and d cannot be formed directly from b. The formation of c and d from the molecular ion a demands a minimum of two and three reactions, respectively. Given that several rearrangements must occur before fragmentation to *m/e* 42 (M − $CDH_2$) is possible, it is reasonable that the relative intensity of *m/e* 42 (M − $CDH_2$) should rise to the statistical value much more slowly than *m/e* 41 (M − $CD_2H$). The rather sharp bend in the *m/e* 42 (M − $CDH_2$) curve in Figure 5b at about $9 \times 10^{-11}$ sec is unexpected and not readily explicable. We note, however, that the corresponding curves (Figure 6b) for 2-methyl-$d_3$-propane-*3,3,3-d*$_3$ are much smoother.

We have shown that reaction Schemes I and II involving 1,3 allylic hydrogen shifts are consistent with the kinetic data. There are reactions other than 1,3 shifts which might conceivably explain the kinetic results. Successive 1,2 shifts are conceivable, but seem unlikely to be responsible for the observed rearrangement. We suggest that 1,2 shifts lead instead to fragmentation and we discuss this possibility below. A second conceivable explanation for the kinetic results involves rearrangement *via* methylcyclopropane-type intermediates. We dismiss this explanation on two grounds. The heat of formation of methylcyclopropane ion is some 30 kcal/mol higher than that of 2-methylpropene ion.[39] Secondly, if methylcyclopropane species were formed, we would have expected them to be detected in the experiments of Lias and Ausloos.[19] The evidence in fact suggests that methylcyclopropane ions isomerize to 2-methylpropene and 2-butene structures. Finally "scrambling" mechanisms have been envisaged to explain H–D randomization.[40] A mechanism (eq 2) of this type would, however, be inconsistent with the kinetic results, since the maximum in the relative intensity of *m/e* 41 (M − $CD_2H$) at short times (Figure 5b) could not be readily explained. The kinetic results are in fact the first good evidence *against* the occurrence of processes in which pairs of hydrogen (deuterium) atoms within a molecule freely exchange positions.



(2)

The kinetic results raise some interesting questions concerning the orbital structure of the reactant radical-ion and the mechanism of its fragmentation to lose methyl. We assume that we can restrict our discussion to the ground electronic state corresponding to ionization of a $\pi$ electron. The possibility exists that the open shell retains the delocalized characteristics of the pristine $\pi$ orbital in the neutral molecule, in which case the 1,3 shift is a sigmatropic reaction. Orbital symmetry considerations[41] demand that 1,3-sigmatropic shifts be antarafacial. Perhaps the 1,3 shift in the 2-methylpropene radical-ion does occur antarafacially. It is, however, unlikely that the analogous 1,3 shift in the cyclohexene radical-ion occurs antarafacially.[18] There are a number of reports of 1,3-sigmatropic rearrangements occurring with stereochemistry forbidden by orbital symmetry.[41b] One explanation[41b,c] offered for the occurrence of such rearrangements is that they are not truly concerted but rather proceed *via* an intermediate. It seems unlikely, however, that an intermediate would be involved in the 1,3 shift in the 2-methylpropene radical-ion. Possibly the 1,3 shift is not a sigmatropic reaction. Perhaps following ionization, the charge is localized on the central carbon to produce a tertiary carbonium ion [see eq 3].



(3)

The odd electron is localized on $C_1$ to produce a primary radical site. Such a planar model structure should be energetically favorable, since the charge can be stabilized by interaction with orbitals of the methyl and methylene groups. In valence bond language the charge is stabilized by hyperconjugation. There would presumably be the same hindered internal rotation about all three C–C bonds. The 1,3 shift now corresponds to a 1,3 radical-induced rearrangement, numerous examples of which are well documented.[12] Orbital symmetry is conserved. There is now the question of why $CH_3$· is eliminated, representing as it does, formally at least, a vinylic cleavage. We suggest that the driving force is

mentation. We can write on the basis of Scheme I

$$[m/e\ 41] = \tfrac{1}{2}[b] \quad \text{(i)}$$

$$[m/e\ 42] = \tfrac{1}{2}[c] + [d] \quad \text{(ii)}$$

$$[m/e\ 43] = [a] + \tfrac{1}{2}[b] + \tfrac{1}{2}[c] \quad \text{(iii)}$$

[*m/e* 41], [*m/e* 42], and [*m/e* 43] are the relative intensities of these ions expressed as *percentages* and can be obtained from Figure 6. [a], [b], [c], and [d] are the *percentage* compositions of the isotopic mixture at $6 \times 10^{-11}$ sec and *for those parent ions with internal energy so as to lose methyl at* $6 \times 10^{-11}$ sec. We can write one more equation

$$\frac{d[m/e\ 41]}{dt} = \tfrac{1}{2}\frac{d[b]}{dt} = \tfrac{1}{2}(6k[a] + k[c] - 3k[b]).$$

At $6 \times 10^{-11}$ sec, when *m/e* 41 (M − $CD_2H$) is at a maximum, it is a good approximation that $d[m/e\ 41]/dt = 0$, so that

$$6[a] + [c] - 3[b] = 0 \quad \text{(iv)}$$

Solving these four equations (eq i–iv) gives [a] = 17%, [b] = 42%, [c] 24%, and [d] = 17%. Thus, at $6 \times 10^{-11}$ sec, some 80% of the molecular ions a have rearranged to b. Some 40% of the 80% have rearranged still further to c and then to d.

(39) J. L. Franklin, J. G. Dillard, H. M. Rosenstock, J. T. Herron, K. Draxl, and F. H. Field, *Nat. Stand. Ref. Data Ser., Nat. Bur. Stand.*, **26**, 260 (1969).

(40) A. N. H. Yeo and D. H. Williams, *J. Amer. Chem. Soc.*, **91**, 3582 (1969).

(41) (a) K. B. Woodward and R. Hoffmann, "The Conservation of Orbital Symmetry," Academic Press, New York, N. Y., 1971, p 114; (b) J. A. Berson, *Accounts Chem. Res.*, **5**, 406 (1972); (c) H. E. Zimmerman, *ibid.*, **5**, 393 (1972); (d) J. E. Baldwin, A. H. Andrist, R. K. Pinschmidt, Jr., *ibid.*, **5**, 402 (1972).

formation of the unusually stable allyl carbonium ion. Convincing support for this suggestion can be marshalled from the appearance potential measurements of Lossing.[42] Formation of the allyl ion implies a 1,2-hydrogen shift [see eq 3]. Such shifts within carbonium ions are well documented.[11] We now have the attractive picture of a highly symmetric planar radical-cation in which hydrogens move rapidly around the peripheral carbons by radical-induced 1,3 shifts. Competing with the radical reaction is a charge-induced 1,2 shift which leads to fragmentation. Admittedly, so far the model is speculative; however, the model contains within it a solid explanation for a baffling discrepancy between measurements by Lossing[42] and certain other measurements by Franklin and colleagues.[43] Lossing, by two separate methods, one being a direct measurement of the ionization potential, estimated the heat of formation of the allyl ion to be 226 kcal/mol. Franklin and colleagues[43] estimated the heat of formation of $C_3H_5^+$ from 2-methylpropene to be 214 kcal/mol by measurements of appearance potential and kinetic energy releases during fragmentation. The disagreement is remarkable, since there is excellent agreement between the two sets of estimations for the heat of formation of $C_3H_5^+$ from 1-butene. It appears that Franklin and coworkers have overcorrected for energy release.[43] Franklin and coworkers assume that the kinetic energy release during fragmentation originates solely from excess energy above the activation energy. This is often a safe assumption for direct bond cleavages, but is rarely valid for rearrangements since these tend to have *reverse* activation energies contributing to kinetic energy release.[44] If loss of methyl from the 2-methylpropene radical-ion is concerted with a 1,2-hydrogen shift, the reaction might be expected to have a reverse activation energy. The measured kinetic energy release would not originate solely from excess energy and use of the formula of Haney and Franklin,[45] $\bar{\epsilon}_t = E^*/\alpha N$, would be unreliable. Under these circumstances it is likely that the correction to the appearance potential would be excessive and would lead to a value for the heat of formation which was too low. We suggest that this is the origin of the discrepancy between the two estimations[42,43] of the heat of formation of $C_3H_5^+$ from 2-methylpropene.

Assuming our mechanistic model to be correct, the results provide information as to the relative activation energies and entropies of a 1,3 radical hydrogen shift as opposed to a 1,2 carbonium ion hydrogen shift. At $6 \times 10^{-11}$ sec, for example, there has on average been between one and two 1,3 shifts prior to a 1,2 shift leading to fragmentation. In general the 1,3 shifts compete more effectively with the 1,2 shifts at longer times. This implies that the 1,3 shift has the lower activation energy but also a lower activation entropy (*vide infra*).

**Reactions Induced by Electron Impact.** The major objective of this part of the discussion is to determine to what extent the reactions induced by FI outside of the very high field (*i.e.*, times a few $\times 10^{-11}$ sec) resemble those induced by EI. The quasiequilibrium theory[36] assumes as its major premise a statistical distribution of internal energy among the internal degrees of freedom regardless of how that energy is imparted to the molecule. If this were actually true, the same reactions would be induced by both FI (at times a few $\times 10^{-11}$ sec) and EI (at the same times); only the relative amounts of the various reactions would differ on account of differing amounts of internal energy. Consider the 16 eV mass spectrum of 2-methylpropene (Figure 7). The major fragment is *m/e* 41 (M − $CH_3$) at an intensity of 10% relative to the molecular ion. The major fragment following FI is also *m/e* 41 (M − $CH_3$). The intensity of the fragment formed between $3 \times 10^{-11}$ and $5 \times 10^{-6}$ sec following FI is 1% relative to the molecular ion. The figure is arrived at by integrating[18] the curve in Figure 8. The higher degree of fragmentation following 16 eV EI can be attributed to a greater amount of excitation energy imparted during the ionization process. The fragment *m/e* 40 (M − $CH_4$) occurs at an intensity of ~1% in the 16 eV EI mass spectrum (too low an intensity to be included in Figure 7) and at an intensity of ~0.1% following FI.

Consider the EI mass spectra of the deuterated compounds (Figure 7). It is evident that considerable rearrangement takes place prior to fragmentation. The fragment intensities for the loss of methyl (Table I) deviate from the values calculated for complete H–D randomization. The intensities of *m/e* 43 (M − $CH_3$) with **1** and *m/e* 44 (M − $CD_3$) with **2** exceed their calculated values, whereas the intensities of *m/e* 42 (M − $CDH_2$) fall below their calculated values. The intensities of *m/e* 41 (M − $CDH_2$) with **1** and *m/e* 46 (M − $CDH_2$) with **2** slightly exceed their calculated values. These deviations from calculated values can be assumed to result from that fragmentation occurring prior to the time when H and D are completely randomized. This being so, we conclude that at times prior to H–D randomization following EI, *m/e* 43 (M − $CH_3$) from **1** and *m/e* 44 (M − $CD_3$) from **2** are formed at intensities in excess of their calculated values, but that *m/e* 42 (M − $CDH_2$) from **1** and *m/e* 45 (M − $CD_2H$) from **2** are formed at intensities less than their calculated values. At these times the fragments *m/e* 41 (M − $CD_2H$) from **1** and *m/e* 46 (M − $CDH_2$) from **2** are formed at intensities slightly in excess of the calculated values. These *deduced* trends for the fragment intensities at times prior to H–D randomization in the case of EI exactly parallel the trends *measured* in the case of FI (Table I). The correspondence between EI and FI for the *m/e* 41 (M − $CDH_2$) from **1** and *m/e* 46 (M − $CDH_2$) from **2** intensities is particularly convincing. We suggest that the reaction Schemes I and II proposed for FI are equally valid for 16 eV and possibly 50 eV EI. We suggest therefore that EI induces rapid 1,3-hydrogen shifts in the same manner as FI. Comparison of the 16 eV and 50 eV EI mass spectra of **1** and **2** (Figure 7) shows that there is less H–D randomization at the higher electron energy (Table I), *i.e.*, fewer 1,3-hydrogen shifts occur prior to fragmentation. Raising the electron energy, therefore, enhances the rate of the fragmentation relative to that of the rearrangement.[46] This would indicate that the 1,3-hydrogen shift has both

(42) F. P. Lossing, *Can. J. Chem.*, **50**, 3973 (1972).
(43) D. K. Sen Sharma and J. L. Franklin, *J. Amer. Chem. Soc.*, **95**, 6562 (1973).
(44) E. G. Jones, J. H. Beynon, and R. G. Cooks, *J. Chem. Phys.*, **57**, 2652 (1972).
(45) M. A. Haney and J. L. Franklin, *J. Chem. Phys.*, **48**, 4093 (1968).
(46) The effects of electron energy on H–D randomization processes have been discussed previously; see A. N. H. Yeo, R. G. Cooks, and D. H. Williams, *Chem. Commun.*, 1269 (1968).

a lower activation energy and a lower activation entropy (or frequency factor[36]) than the fragmentation to lose methyl.[47] The same conclusion can be established[18] somewhat more securely from, and is implicit in, our discussion of the FIK results (Figures 5 and 6).

It has been common practice among mass spectroscopists to refer to all such H–D rearrangement as we have been discussing as "scrambling." The term "scrambling" seems to hold a mechanistic implication, namely that H and D atoms move and exchange positions in a random manner. This is not the case with the 2-methylpropene ion. We suggest, mainly on the basis of FIK results,[17,18,47–49] that with all functionalized aliphatic ions the predominant reactions leading to H–D randomization are *specific* and *definable hydrogen shifts*. The nature of the shifts will depend very much on the nature of the molecule; *e.g.*, shifts in alkene ions[17,18] bear little relationship to those in ketone ions.[48] In such a situation the term "scrambling" is seriously and objectionably misleading. The mechanistic implication is false and there is the false suggestion of some common characteristics or links among processes which are actually widely disparate. We suggest therefore that the use of the term "scrambling" be discontinued. In those cases where the rearrangements cannot be identified more precisely, it seems sufficient to refer to "rearrangements leading to H–D randomization."

(47) For a discussion of the usefulness of the concepts of energy and entropy of activation in understanding unimolecular reactions of (radical-) cations and in particular FIK results, see P. J. Derrick and A. L. Burlingame, *Accounts Chem. Res.*, in press.

(48) P. J. Derrick, A. M. Falick, A. L. Burlingame, and C. Djerassi, *J. Amer. Chem. Soc.*, **96**, 1054 (1974).

(49) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Chem. Soc., Perkin Trans. 2*, submitted for publication.

## Conclusion

The study of this relatively simple system establishes with some degree of certainty that 1,3 allylic hydrogen shifts are facile and rapid unimolecular gas-phase reactions of the 2-methylpropene radical-cation. This is an important result in itself. Moreover, similar 1,3 allylic hydrogen shifts must be expected to occur in other alkene radical-cations. This knowledge will be a considerable asset in unraveling the complexities of the chemistry of such reaction systems. Further FIK experiments are in progress with butenes and other alkenes.

**Acknowledgment.** We are indebted to the National Science Foundation (Grant GP 38389X) and the National Aeronautics and Space Administration (Grant NGL 05-003-003) for financial support. We thank Miss Rosalynd Jackson for measuring the EI mass spectra.

# DEUTERIUM AND CARBON-13 TRACER STUDIES OF ETHANOL METABOLISM IN THE RAT BY $^{2}H,^{1}H$-DECOUPLED $^{13}C$ NUCLEAR MAGNETIC RESONANCE

D. M. Wilson and A. L. Burlingame
Space Sciences Laboratory,
University of California, Berkeley, California 94720

T. Cronholm and J. Sjövall
Kemiska Institutionen,
Karolinska Institutet, Stockholm 60, Sweden

Received December 17,1973

SUMMARY: [1-$^{13}C$, 1,1-$^{2}H_2$] ethanol and [2,2,2-$^{2}H_3$] ethanol were administered to bile fistula rats. A new technique, $^{2}H,^{1}H$-decoupled $^{13}C$ nuclear magnetic resonance, was used in attempting to account for the distribution of the isotopic species along the steroid skeleton of 3-45 mg of isolated bile acids. The technique revealed $^{2}H$ incorporation at many carbon sites unambiguously, but has limitations as a quantitative $^{2}H$ assay at these levels of sample availability.

Interest in charting metabolic pathways by non-radioactive isotope incorporation in vivo has resulted in the exploration of new instrumental techniques to overcome the many orders of magnitude difference in detectability of $^{2}H$, $^{13}C$, $^{17}O$, in comparison to their radioactive counterparts. $^{2}H$-incorporation into metabolic products on the sub-nanogram level can be detected by the molecular ion m/e ratios in mass spectra (1), but the site of incorporation in the molecule is often unattainable by this technique. At larger sample levels, $^{1}H$ or $^{2}H$ nuclear magnetic resonance (nmr) may be expected to reveal incorporation sites, but only in relatively simple cases where attribution of individual resonances to specific nuclei is possible. The greatly increased resolution of $^{1}H$-decoupled $^{13}C$ nmr has recently been shown to allow $^{2}H$-site detection through $^{2}H$-$^{13}C$ isotopic shifts and coupling effects (2). We wish to report here a further improvement in this technique, that of simultaneous $^{2}H$- and $^{1}H$-decoupled $^{13}C$ nmr, to detect in vivo incorporation of $^{2}H$ from [1-$^{13}C$, 1,1-$^{2}H_2$] ethanol and [2,2,2-$^{2}H_3$] ethanol in bile acids of the rat. In a study of ethanol metabolism, this biosynthetic system has the advantage that labeled atoms are introduced in two separate pathways, i.e., via acetate and via reduced pyridine nucleotides. Furthermore, a relatively large quantity of product for analysis is obtained from a single animal. Parallel studies by gas chromatography-mass spectrometry are reported elsewhere (3).

## MATERIALS AND METHODS

[1-$^{13}C$, 1,1-$^{2}H_2$] ethanol and [2,2,2-$^{2}H_3$] ethanol were synthesized from



Copyright © 1974 by Academic Press, Inc.
All rights of reproduction in any form reserved.
REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

nominally 60% $^{13}C$-enriched acetic acid, and 95% $^{2}H$-enriched [2,2,2-$^{2}H_3$] acetic acid, respectively, and analyzed by gas chromatography-mass spectrometry of the benzoate. Male Sprague-Dawley rats weighing about 200 g were provided with bile fistulae and thin catheters for intraperitoneal injections. The labeled ethanol was administered in small aliquots (in saline) starting 48 hr after the operation, and concentration of ethanol in bile was monitored by gas chromatography. The volumes injected were adjusted to maintain a concentration of ethanol above 10 mM. The [1-$^{13}C$, 1,1-$^{2}H_2$] ethanol was administered to one rat and bile collected for 23 hr. The sample taken between 13 and 18 hr was used for analysis. [2,2,2-$^{2}H_3$] ethanol was administered to two rats and bile collected for 48 hr. The rats were kept in restraining cages and had free access to food pellets and saline.

Bile acids were extracted, saponified, methylated, and separated on hydroxyalkoxypropyl Sephadex or silicic acid (3,4). Aliquots were converted into trimethylsilyl ethers and heavy isotope content was determined using a 1% SE-30 column at 240°C in an LKB 9000 gas chromatography-mass spectrometry instrument, in conjunction with an incremental magnetic tape spectral recording system (5). Background subtraction, normalization, plotting and calculation of heavy isotope abundance was performed on an IBM 1800 computer (6). Methyl [3β-$^{2}H$] cholate, [3β,7β-$^{2}H_2$] cholate, [3β-$^{2}H$] chenodeoxycholate and [7β-$^{2}H$] chenodeoxycholate were synthesized by $NaB^{2}H_4$ reduction of the corresponding oxo-esters, for use as nmr standards. Other bile acid methyl esters were obtained commercially to assist in assigning $^{13}C$ nmr resonances.

A Varian XL-100-15 $^{13}C$ Fourier transform nmr spectrometer in conjunction with an XDS Sigma 7 computer system (7) was used to record the $^{13}C$ spectra. The nmr probe was fitted with an "external" pulse-modulated lock, which included a $Cu^{++}$-doped LiCl capillary, transmitter, preamplifier and 38.929 MHz phase locked crystal oscillator. A 15.361 MHz frequency modulated crystal oscillator and 10 w amplifier, coupled to the $^{13}C$ transmitter coils through a matching network, comprised the $^{2}H$-decoupling system. Bile acid methyl ester samples (3-45 mg) were dissolved in 5 mm nmr tubes, and spectra recorded at 50°C using radio-frequency pulses placed in the center of the $^{13}C$ spectrum at 5 sec intervals. A 200 μsec A/D sampling rate and dual-phase detection (8) in conjunction with a 32-bit 16,384-word computer memory block afforded $^{13}C$ shift resolution of ±0.02 ppm, after 2000-15000 accumulated free-induction decays per channel and subsequent Fourier transform. A trace of tetramethylsilane was added to all samples to serve as an internal shift reference. The shift reagent Eu(FOD)$_3$ (Bio-Rad Laboratories) was used in chloroform solvent in some cases to assist in confirming $^{13}C$ spectral assignments.

## RESULTS

$^{13}C$ chemical shift assignments of unenriched methyl cholate and chenodeoxycholate in pyridine were made based upon comparisons of their spectra with those of deuterated or acetylated analogs, spectra of lithocholate and deoxycholate, known biosynthetic routes for carbon incorporation, LIS-reagent studies, published assignments (9) of other steroids, and off-resonance $^{1}H$-decoupled spectra. Similar spectra in chloroform and 50% (v/v) pyridine-chloroform demonstrated a "scrambling" effect on the order of these assignments in their appearance from low to high fields, as has been reported elsewhere (10) for cholesterol, pyridine, and uridine. For example, in chloroform the $C_{23}$ of methyl cholate shifts downfield by 0.8 ppm (when corrected for susceptibility differences and internal reference solvent dependencies) from its position in pyridine, while $C_2$ crosses it and shifts upfield by 0.4 ppm. The expectation that these spectral shifts would be helpful in uncovering $^{13}C$-$^{2}H$ isotope peaks hidden by nearby resonances was not realized, however, because of line broadening in chloroform due to aggregation (11). For the same reason, LIS reagent-induced shifts were not used in the remainder of this work, although a marked binding preference of the chelate to the $C_3$ position in bile acids was noted. At the time this manuscript was nearing completion, the $^{13}C$ spectral assignments of bile acids were published (12) which were identical to our results, if consideration is given to the solvent effects noted; therefore our assignments will not be given here.

Mass spectrometric analysis of methyl cholate and chenodeoxycholate derived from [1-$^{13}C$, 1,1-$^{2}H_2$] ethanol administration showed an excess of about 220 atoms of heavy isotope per 100 molecules. Results of previous work (3) using [1,1-$^{2}H_2$] ethanol indicated that about 60 of these atoms were the $^{2}H$ isotope. About 40% of the molecules were entirely unlabeled. Analysis of methyl cholate derived from [2,2,2-$^{2}H_3$] ethanol administration showed a deuterium excess of 280 atoms per 100 molecules. If it is assumed that deuterium is not lost from $C_2$ of ethanol during cholesterol biosynthesis this corresponds to 17 atoms percent excess in each position expected to be labeled. About 45% of the molecules were unlabeled.

Heavy isotope incorporation at specific carbon sites in rat bile acids, as determined by $^{2}H$,$^{1}H$-decoupled $^{13}C$ nmr is shown in Table I. Portions of comparative spectra which allowed these measurements are shown in Fig. 1.

## DISCUSSION

Evidence is growing that at least for tetrahedral carbon, $^{13}C$-$^{2}H$ isotope shifts are predictable, i.e., upfield, and linearly related to the number of directly attached, geminal, and vicinal deuterons (13). The $^{2}H$,$^{1}H$-decoupled

TABLE I

Heavy Isotope Incorporation in Bile Acids[a]

| Carbon | $^{13}C$ enrichment[b] | | $^{2}H$ incorporation (%)[d], & $^{2}H$ isotope shift (ppm)[e] | | |
|---|---|---|---|---|---|
| | 1a | 2a | 1a | 2a | 1b |
| 1 | | | | | δ |
| 2 | c | 13 | δ | 9, .40 | δ |
| 3 | | | | | 0 |
| 4 | 19 | 18 | δ | 0 | δ |
| 5 | | | | | δ |
| 6 | c | 20 | δ | δ | δ |
| 7 | | | | | 8, .47 |
| 8[k] | 19 | 18 | 0 | 0 | 0 |
| 9 | | | | | 0 |
| 10[g] | 18 | c | | | |
| 11[j] | 14 | 17 | 8, .36 | 9, .35 | 0 |
| 12[j] | 17 | 18 | δ | 10, .36 | δ |
| 13[g] | | | | | |
| 14[k] | 16 | 19 | 0 | 0 | 0 |
| 15 | | | | | 9, .34 |
| 16 | 16 | 21 | 12, .36 | 13, .39 | δ |
| 17 | | | | | 6, .47 |
| 18 | | | | | (12, .22)[h] |
| 19 | | | | | (16, .19)[i] |
| 20 | c | c | δ | 0 | δ |
| 21 | | | | | 6, .86 |
| 22 | | | | | δ |
| 23 | 24 | 26 | 0 | 3, .27 | 0 |
| 24[g] | | | | | |

[a] 1a and 2a are methyl cholate and methyl chenodeoxycholate, respectively, from bile collected between 13-18 hr after the beginning of [1-$^{13}C$, 1,1-$^{2}H_2$] ethanol administration. 1b is methyl cholate from bile collected between 24-48 hr after the beginning of [2,2,2-$^{2}H_3$] ethanol administration. The quantities available for nmr analyses were: 1a and 2a, approximately 3 mg each; 1b, 45 mg.

[b] in multiples of natural abundance, assuming no labeling of carbon atoms derived from $C_2$ of acetate. To avoid large errors resulting from unequal nuclear Overhauser enhancement (N.O.E.) and spin-lattice relaxation times ($T_1$), the weak natural abundance $^{13}C$ resonances of 1a and 2a were integrated, and compared with the integrated intensity of the same resonances of standards observed under identical experimental conditions, resulting in normalization

factors $S_{1a}$ and $S_{2a}$. The ratios of $^{13}C$-enriched intensities in 1a and 2a to their same intensities in the standard spectra were then multiplied by these normalization factors to arrive at the enrichment multiples. Thus, although the absolute values of enrichment are subject to the high experimental uncertainty (±25%) in the natural abundance intensities of 1a and 2a, the relative enrichment of $^{13}C$ sites is valid to ca. ±5%. Small errors resulting from new N.O.E.'s and $T_1$'s due to partial $^2H$ incorporation are ignored.

[c] accurate calculation not possible due to overlapping peaks in $^{13}C$ enriched or standard spectra.

[d] minimum percentage (see discussion) of $^{13}C$ atoms carrying $^2H$ atom(s); calculated as the difference in integrated areas of $^2H$-shifted peaks between comparable $^2H$, $^1H$-decoupled and $^1H$-decoupled spectra, divided by the integrated $^{13}C$ present at that site.

[e] all isotope shifts are to higher field.

[f] unobservable due to peak overlap.

[g] non-protonated carbon.

[h] geminal isotope shift of $C_{13}$.

[i] geminal isotope shift of $C_{10}$.

[j] $J_{C_{11}C_{12}}$ of 1a, 36.4 Hz; of 2a, 32.4 Hz.

[k] $J_{C_8C_{14}}$ of 1a, 35.6 Hz; of 2a, 36.4 Hz.

---

technique is designed to observe the largest of these, i.e., the carbon to which the deuteron is directly attached, so it should be able to detect partial deuteration in methylene and methyl groups quite easily. In fact, in Table I all the methylene groups demonstrating $^2H$-incorporation are believed to contain only a single deuteron for this reason.

The most appealing aspect of the $^2H$-incorporation detection method described here is that a change in the setting of the $^2H$ decoupler (i.e., "switching" the $J_{^{13}C^2H}$-coupling on or off) is all that is required to observe the qualitative effect, without additional chemical or instrumental procedures. However, the quantitative measure of $^2H$-incorporation at specific carbon sites by our "difference" method (see footnote d, Table I) results in systematically low values since some isotopically shifted intensity is present in $^2H$-coupled spectra. Other factors also lead to systematic errors in this direction. Allerhand et al. (14) have reported that carbons in cholesteryl chloride relax mainly through $^{13}C$-$^1H$ dipolar interactions. Since $^{13}C$-$^2H$ dipolar relaxation is approximately 16 times less efficient, one would expect the relaxation times and $^1H$ nuclear Overhauser enhancements of individual carbons in partially deuterated molecules to depend on proximity to nearby protons in a complicated manner. In this study, we have not used instrumental or paramagnetic ion methods to suppress nuclear Overhauser enhancement (15,16) or reduce carbon $T_1$'s (16), but in any case the $^2H$,$^1H$-decoupled $^{13}C$ nmr technique can produce only the lower estimate of $^2H$ site incorporation. For example, if the non-

deuterated carbon $T_1$'s of $C_{18}$ and $C_{19}$ in methyl cholate are nearly equal to 1.5 sec as in cholesteryl chloride (14), complete deuteration as might be expected in the case of [2,2,2-$^2H_3$] ethanol administration would render these carbons invisible, under the conditions our data were taken, since the nearest protons are 3 bonds removed. This was in fact the case, indicating that there is apparently no exchange of deuterium at ethanol-$C_2$ on oxidation to acetate and activation to acetyl CoA. The $^2H$ incorporation at $C_{18}$ and $C_{19}$ was instead detected in the geminal $^{13}C$-$^2H$ isotope shifts experienced by $C_{10}$ and $C_{13}$.

Estimates of full or partial deuterium content should also be gained from the decrease in intensities of residual undeuterated peaks in $^1H$-decoupled $^{13}C$ spectra, from their values in spectra of unenriched standards run under identical conditions. This method would be expected to give systematically low estimates since it ignores overlapping of $^2H$-coupled peaks (see also ref. 2), but particularly fails here because it depends on small differences in large numbers (intensities) with relatively large uncertainties. However, when applied to the spectra of methyl cholate derived from [2,2,2-$^2H_3$] ethanol administration, the largest decreases in intensity occurred in the $C_7$, $C_{15}$, $C_{17}$, $C_{18}$ and $C_{21}$ resonances, and suggested that ~15% of each of these carbon atoms was deuterated.

Double ($^{13}C$,$^2H$) isotopic incorporation in bile acids from [1-$^{13}C$, 1,1-$^2H_2$] ethanol results in marked $^{13}C$ incorporation at 11 carbon sites. If it is assumed that carbon atoms derived from $C_2$ of acetate are unlabeled, the total $^{13}C$ excess of the molecules can be calculated from the nmr data to be about 210 atoms percent. Mass spectrometric analysis indicated 160 atoms percent excess. This supports the contention that only carbon atoms derived from $C_1$ of ethanol are labeled. Deuterium is transferred from ethanol to the 4-pro-R site of NADH and would be detected by our technique only if it were incorporated at one of these carbon atoms after transfer to the NADPH used as cofactor in the biosynthesis. The primary reduction of hydroxymethyl glutaryl-CoA to mevalonate resulted in predictable $^2H$ incorporation at some of these sites. However, $^2H$-incorporation at $C_{14}$ is conspicuously absent, and suggests that the 4-pro-S site of NADPH used in the $C_{14}$ demethylation of lanosterol (17) was unlabeled. The low deuterium content at $C_{23}$ is explained by partial loss during alkaline hydrolysis of the conjugated bile acids.

Since the cyclization of squalene to lanosterol is believed to be symmetric, it is difficult to account for the relative differences in $^{13}C$ enrichment between $C_2$ and $C_{23}$ (Table I, also see Fig. 1). A reduction in integrated intensity of the undeuterated $C_2$ peak not accounted for by a corresponding increase in its $^2H$-isotopically shifted resonance because of nuclear Overhauser or $T_1$ effects is not expected as it is predominantly singly deuterated as evi-

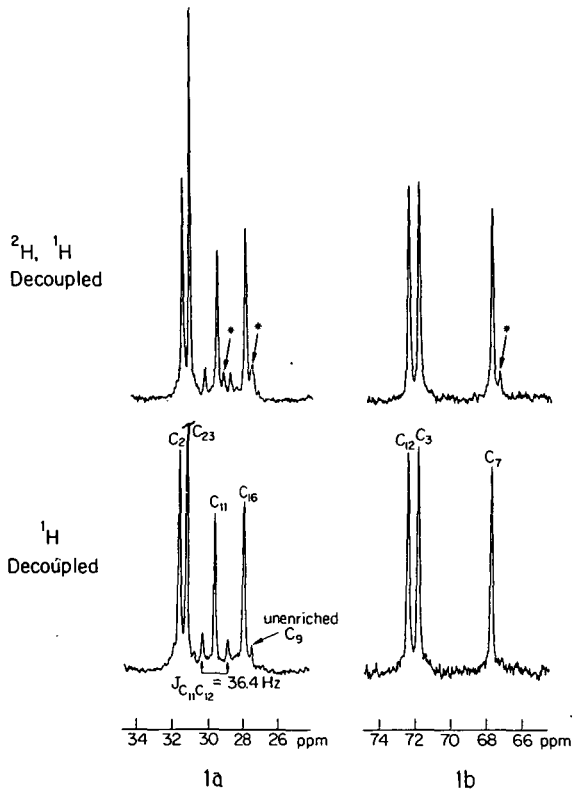



Figure 1. Portions of $^{13}C$ spectra of (1a) methyl cholate derived from [1-$^{13}C$, 1,1-$^2H_2$] ethanol administration, and (1b) methyl cholate derived from [2,2,2-$^2H_3$] ethanol administration. Asterisks identify resonances of deuterated carbons, which are separated from their protonated parent peaks at lower field by $^{13}C$-$^2H$ isotope shifts. The solvent was pyridine at 50°C. The chemical shift scale is in ppm downfield from internal TMS.

denced by its isotope shift. For the same reason, a large increase in total integrated intensity at $C_{23}$, because of partial deuterium loss on alkaline hydrolysis, would not be expected. We shortly plan to complement the results given here by investigation of the other 13 carbon sites using [2-$^{13}C$, 1,1-$^2H$] ethanol, which may provide some insight into this question.

ACKNOWLEDGEMENT: Financial support from the National Aeronautics and Space Administration Grant NGL 05-003-003 and the Swedish Medical Research Council Grant 13X-2189 is gratefully acknowledged; the authors thank Dr. Peter Christiansen and Ms. Sydell Lewis for the synthesis of labeled ethanol, Miss Birgitta Lindskog for laboratory work and Mr. R. W. Olsen for instrumental design and modification assistance in the performance of this research. One of us (A.L.B.) is grateful to the John Simon Guggenheim Memorial Foundation for a fellowship, 1970-1972.

REFERENCES

1. For reviews of the application of mass spectrometry to the analysis of

stable isotopic incorporation, see Ratner, S. (1972) in Biochemical Applications of Mass Spectrometry (Waller, G. R., Ed.), pp. 1-10, Wiley-Interscience, New York; Caprioli, R. M. (1972) ibid., 735-776; also Grostic, M. F., and Rinehart, K. L., Jr. (1971) in Mass Spectrometry: Techniques and Applications (Milne, G. W. A., Ed.), pp. 217-287, Wiley-Interscience, New York.

2. Stothers, J. B., and Tan, C. T. (1972) J. Amer. Chem. Soc. 94, 8581-8582.
3. Cronholm, T., Burlingame, A. L., and Sjövall, J. (1973), J. Biol. Chem., submitted for publication; also Cronholm, T., Burlingame, A. L., and Sjövall, J. (1973) in Abstracts, Ninth International Congress of Biochemistry, Stockholm, 1-7 July, p. 412.
4. Cronholm, T., Makino, I., and Sjövall, J. (1972) European J. Biochem. 24, 507-519.
5. Hedfjäll, B., Jansson, P.-A., Mårde, Y., Ryhage, R., and Wikström, S. (1969) J. Sci. Instr. 2, 1031-1035.
6. Reimendal, R., and Sjövall, J. (1971) in Hormonal Steroids (James, V.H.T. and Martini, L., Eds.), pp. 228-237, Excerpta Medica, ICS 219, Amsterdam.
7. McPherron, R., Olsen, R. W., and Burlingame, A. L. (1973) in Advances in Mass Spectrometry (West, A. R., Ed.), Institute of Petroleum, London, Vol. 6, in press.
8. Wilson, D. M., Olsen, R. W., and Burlingame, A. L., Rev. Sci. Instr., submitted for publication.
9. Reich, H. J., Jautelat, M., Messe, M. T., Weigert, F. J., and Roberts, J. D. (1969) J. Amer. Chem. Soc. 91, 7445-7454.
10. Mantsch, H. H., and Smith, I. C. P. (1973) Canad. J. Chem. 51, 1384-1391.
11. Bennet, W. S., Eglinton, G., and Kovac, S. (1967) Nature 214, 776-780.
12. Leibfritz, D., and Roberts, J. D. (1973) J. Amer. Chem. Soc. 95, 4996-5003.
13. Colli, H. N., Gold, V., and Pearson, J. E. (1973) J. C. S. Chem. Comm., 408-409, and references therein; Breitmaier, E., Jung, G., Voelter, W., and Pohl, L. (1973) Tetrahedron 29, 2485-2489.
14. Allerhand, A., Doddrell, D., and Komoroski, R. (1971) J. Chem. Phys. 55, 189-198.
15. Freeman, R., Hill, H. D. W., and Kaptein, R. (1972) J. Magnetic Resonance 7, 327-329.
16. Freeman, R., Pachler, K. G. R., and La Mar, G. N. (1971) J. Chem. Phys. 55, 4586-4593; Gansow, O., and Burke, A. (1972) Chem. & Eng. News, April 10, 15.
17. Wilton, D. C., Watkinson, I. A., and Akhtar, M. (1970) Biochem. J. 119, 673-675.

[Reprinted from the Journal of the American Chemical Society, 96, 1054 (1974).]
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner.

REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

# Hydrogen–Deuterium Randomization in 7-Methyl-4-octanone-7-$d_1$ at Times of $10^{-11}$–$10^{-5}$ Sec Following Field Ionization

**P. J. Derrick,*[1a,b] A. M. Falick,[1a] A. L. Burlingame,[1a] and Carl Djerassi[1c]**

*Contribution from Space Sciences Laboratory, University of California, Berkeley, California 94720, and Department of Chemistry, Stanford University, Stanford, California 94305.* *Received July 14, 1973*

**Abstract:** Partial hydrogen–deuterium (H–D) randomization occurs prior to the McLafferty rearrangement in 7-methyl-4-octanone-7-$d_1$ following field ionization (FI) even at times as short as the order of $10^{-11}$ sec. This behavior contrasts sharply with that of straight-chain ketones in which no H–D randomization occurs prior to the McLafferty rearrangement at times $<7 \times 10^{-10}$ sec. At times of $10^{-11}$–$10^{-10}$ sec following FI the rate of the McLafferty rearrangement is an order of magnitude smaller in 7-methyl-4-octanone than in a typical straight-chain ketone 2-octanone, but at $\sim 10^{-6}$ sec the situation is reversed with the rate being an order of magnitude smaller in the straight-chain ketone. The difference in kinetics is consistent with the McLafferty rearrangement at times > a few $\times 10^{-11}$ sec following FI being a stepwise rather than a concerted process. It is proposed that H–D randomization in the 7-methyl-4-octanone-7-$d_1$ ion involves $\gamma$-D transfer to the oxygen forming a tertiary radical, followed by reverse D transfer from the oxygen to the alkyl chain. The reverse transfer is facilitated by the stability of the tertiary radical.

The unimolecular gas-phase reactions induced by electron impact (EI) of aliphatic ketones have been studied intensively.[2] The behavior of aliphatic ketones following field ionization (FI) has also received considerable attention.[3–6] Mass spectra of partially deuterated straight-chain aliphatic ketones produced by EI at ordinary ionizing voltages (70 eV) reveal little hydrogen–deuterium (H–D) randomization,[7] although a degree of randomization is manifested in the metastable region[8] (corresponding to molecular ion lifetimes of $10^{-6}$ sec). At very low ionizing voltages (10–12 eV) a degree of H–D randomization is also evident in the normal EI mass spectra.[7,8] These observations have been rationalized on the basis that the randomization reactions in straight-chain aliphatic ketones have low

(1) (a) University of California; (b) Department of Chemistry, University College, University of London; (c) Stanford University.

(2) For a review of the early studies, see H. Budzikiewicz, C. Djerassi, and D. H. Williams, "Mass Spectrometry of Organic Compounds," Holden-Day, San Francisco, Calif., 1967, Chapter 3.

(3) A. J. B. Robertson and B. W. Viney, *Advan. Mass Spectrom.*, 3, 23 (1966).

(4) F. W. Röllgen and H. D. Beckey, *Surface Sci.*, 23, 69 (1971).

(5) E. M. Chait and F. G. Kitson, *Org. Mass Spectrom.*, 3, 533 (1970).

(6) P. J. Derrick, A. M. Falick, S. Lewis, and A. L. Burlingame, *Org. Mass Spectrom.*, 7, 887 (1973).

(7) W. Carpenter, A. M. Duffield, and C. Djerassi, *J. Amer. Chem. Soc.*, 90, 160 (1968).

(8) (a) A. N. H. Yeo, R. G. Cooks, and D. H. Williams, *Chem. Commun.*, 1269 (1968); (b) A. N. H. Yeo and D. H. Williams, *J. Amer. Chem. Soc.*, 91, 3582 (1969).

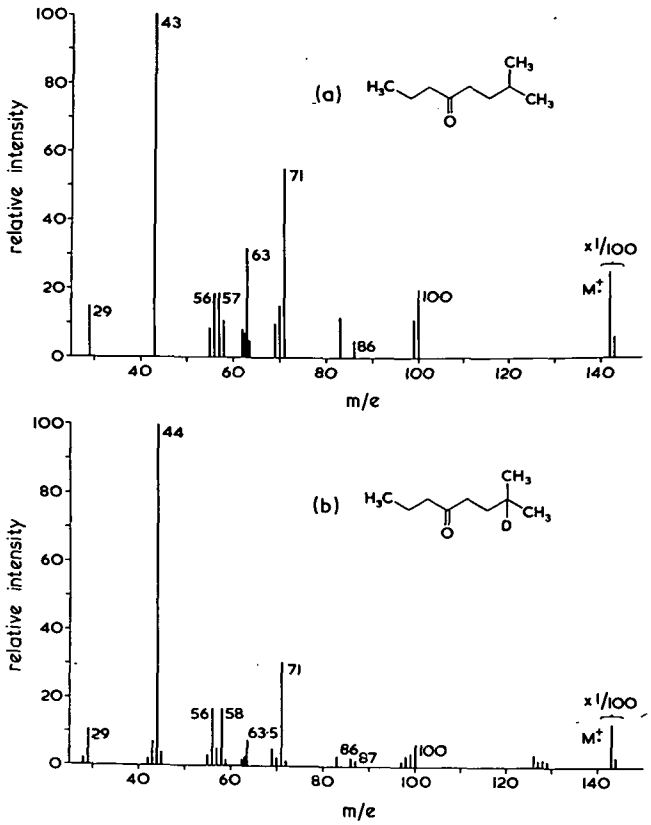



Figure 1. The "normal" FI mass spectrum of (a) 7-methyl-4-octanone; (b) 7-methyl-4-octanone-7-$d_1$. Source temperature 50°. Uncoated Schick blade emitter.

activation energies and frequency factors[8] (employing the concepts of the simplified quasi-equilibrium theory[9]). A study[6] of H–D randomization in 2-octanone-*1,1,1,3,3-*$d_5$ following FI supports this view. No H–D randomization occurs prior to the McLafferty rearrangement at times less than $7 \times 10^{-10}$ sec following FI. Partial H–D randomization is observed at times of $10^{-6}$–$10^{-5}$ sec. The EI mass spectra of certain branched aliphatic ketones, however, contain evidence of extensive H–D randomization even at ordinary ionizing voltages.[10] Compounds labeled specifically at the $\gamma$ position seem to exhibit more randomization prior to the McLafferty rearrangement than compounds labeled elsewhere.[10] Branching at the $\gamma$ position would seem to be more conducive to randomization prior to the McLafferty rearrangement than branching at the $\delta$ position.[10]

We have sought to elucidate this difference in behavior between certain branched and straight-chain ketones as regards H–D randomization by studying the randomization following FI in the $\gamma$-branched and $\gamma$-d-labeled 7-methyl-4-octanone-7-$d_1$. FI mass spectrometry permits rates of unimolecular gas-phase reactions to be measured at times as short as $10^{-11}$ sec (or even $10^{-12}$ sec) and at longer times.[11–16] We shall refer to

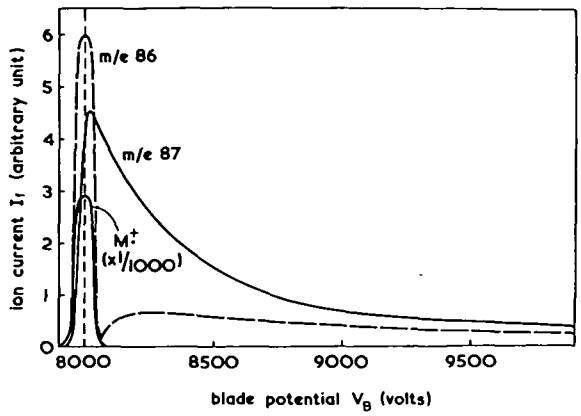



Figure 2. The experimental curves of ion currents $I_f$ against blade potential $V_B$ for the molecular ion and the McLafferty rearrangement products from 7-methyl-4-octanone-7-$d_1$. Each curve is the average of several runs. Considering, for example, *m/e* 86 between $V_B$ = 8100 V and $V_B$ = 9900 V, fluctuations in any run are of the order of 10% of the signal. Fluctuation is proportionately less for higher ion currents.

such data on rates of reaction from times of $10^{-11}$ to $10^{-5}$ sec as "field ionization kinetics (FIK)."[17] The highly resolved (according to time) view of events afforded by FIK can often allow a deeper and more certain understanding of the reactions responsible for H–D randomization than is possible from the integrated view (over some $10^{-6}$ sec) afforded by EI alone.[17,18]

## Results

The "normal" FI mass spectra of 7-methyl-4-octanone and 7-methyl-4-octanone-7-$d_1$ obtained with a double focusing mass spectrometer are reproduced in Figure 1. EI mass spectra at 70 and 15 eV ionizing energies have been published.[10] "Normal" FI mass spectra, as defined in this paper, are those measured with the potential $V_B$ applied to the blade emitter approximately equal to the accelerating potential (8000 V) necessary for ions to be transmitted through the electric sector analyzer. Thus the "normal" spectra display the ions formed within the very high field in the immediate vicinity of the blade (*i.e.*, at times less than $\sim 10^{-11}$ sec).[16]

Profiles of ion current against blade potential $V_B$ for the molecular ion and the important fragment ions following FI of 7-methyl-4-octanone-7-$d_1$ are reproduced in Figure 2. The profiles are measured on an XY recorder by focusing the magnetic analyzer on a particular ion and sweeping the blade potential.[16] The theory behind the measurement and interpretation of such profiles has been discussed in detail elsewhere (see, for example, ref 14, 16, and 17). The measured profiles of ion current $I_f$ against blade potential for the fragment ions can be transformed to curves of ion current $I_f(t)$ against molecular ion lifetime[16,19] (Figure 3). The fragments of *m/e* 86 ($C_5H_{10}O$) and *m/e* 87 ($C_5H_9DO$) are

(9) H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig, and H. Eyring, *Proc. Nat. Acad. Sci. U. S.*, 38, 667 (1952).
(10) G. Eadon and C. Djerassi, *J. Amer. Chem. Soc.*, 92, 3084 (1970).
(11) H. D. Beckey, *Z. Naturforsch. A*, 16, 505 (1961).
(12) H. D. Beckey and H. Knöppel, *Z. Naturforsch. A*, 21, 1920 (1966).
(13) H. D. Beckey, H. Hey, K. Levsen, and G. Tenschert, *Int. J. Mass Spectrom. Ion Phys.*, 2, 101 (1969).
(14) P. J. Derrick and A. J. B. Robertson, *Proc. Roy. Soc., Ser. A*, 324, 491 (1971).
(15) K. Levsen and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, 7, 341 (1971).
(16) A. M. Falick, P. J. Derrick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, 12, 101 (1973).
(17) P. J. Derrick, A. M. Falick, and A. L. Burlingame, 20th Annual Conference on Mass Spectrometry and Allied Topics, Dallas, Texas, 1972, Paper No. H4.
(18) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, 94, 6794 (1972).
(19) J.-P. Pfeifer, A. M. Falick, and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.*, 11, 345 (1973).

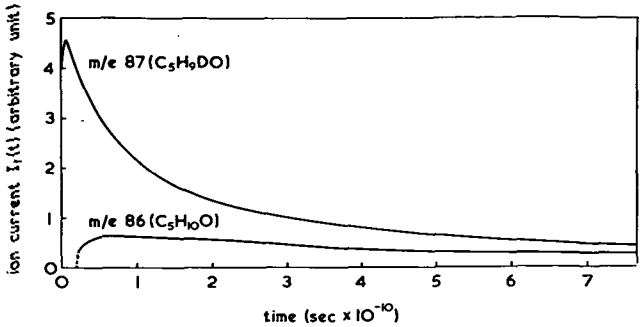



**Figure 3.** Ion currents $I_f(t)$ of *m/e* 86 ($C_5H_{10}O$) and *m/e* 87 ($C_5H_9DO$) from 7-methyl-4-octanone-7-$d_1$ as a function of molecular ion lifetime. The times are derived by mathematical procedure[16] from the blade potentials $V_B$ in Figure 2.

products of the McLafferty rearrangement (see below). The portion of the *m/e* 86 profile at $V_B < 8070$ V in Figure 2 has been disregarded in drawing up Figure 3, since this has probably arisen from a surface process.[20] The ratio of the ion currents $I_f(t)$ for *m/e* 86 and 87 at any particular time is approximately equal to the ratio of their rates of formation at that time.[16,18]

Phenomenological rate constants $\bar{k}(t)$ can be calculated for the formation of a fragment ion from the molecular ion from eq 1 which has been discussed

$$\bar{k}(t) = I_f(t)/I_M\Delta t \qquad (1)$$

fully.[16,18,21] A phenomenological rate constant $\bar{k}(t)$ is an experimental quantity referring to a real situation in which reactant ions exist over a distribution of internal energies. A phenomenological rate constant is distinct from a microscopic rate constant, which is the theoretical quantity referring to the ideal situation in which reactant ions possess some discrete amount of internal energy. Phenomenological rate constants are function of time[16] and are therefore calculated at various reaction times $t$. $I_f(t)$ is the fragment ion current at any time $t$ and $I_M$ represents the ion current of molecular ion available for reaction at that time. $\Delta t$ accounts for the fact that fragment ions formed over a small time interval contribute to the measured ion current $I_f(t)$.[16] Phenomenological rate constants for decompositions in the first and second field free region can be estimated by substituting metastable intensities into eq 1 for $I_f(t)$. In Table I the phenomenological rate constants $\bar{k}(t)$ calculated from eq 1 for the formation of *m/e* 86 ($C_5H_{10}O$) from 7-methyl-4-octanone by the McLafferty rearrangement following FI are compared with rate constants calculated in the same way for the McLafferty rearrangement following FI of a typical straight-chain ketone 2-octanone.

The relative intensities of the metastable transitions $143^+ \rightarrow 87^+$ and $143^+ \rightarrow 86^+$ induced by FI of 7-methyl-4-octanone-7-$d_1$ in the first and second field free regions are shown in Table II. The metastables in the first field free region are measured by defocusing the electric sector analyzer in the normal manner.[22] The figures for the shorter times following FI are taken from Figure 3. Intensities of *m/e* 86 and 87 in the EI mass spectra measured by earlier workers[10] are also shown in Table II.

**Table I.** Relative Phenomenological Rate Constants $\bar{k}(t)$[a] for the McLafferty Rearrangement Following FI in 7-Methyl-4-octanone[b] and in 2-Octanone[c] (Given Relative to that for the McLafferty Rearrangement in 7-Methyl-4-octanone at $1 \times 10^{-5}$ sec)

| Time, sec | 7-Methyl-4-octanone | 2-Octanone |
|---|---|---|
| $7 \times 10^{-11}$ | $1 \times 10^5$ | $8 \times 10^5$ |
| $7 \times 10^{-10}$ | $8 \times 10^3$ | $6 \times 10^4$ |
| $2 \times 10^{-6}$ | 8 | 1.4 |
| $1 \times 10^{-5}$ | 1 | Undetected |

[a] Maximum possible error ±25%. [b] Measurements made at 50°. [c] Measurements made at 25°.

**Table II.** Relative Intensities of *m/e* 87 ($C_5H_9DO$) and *m/e* 86 ($C_5H_{10}O$) from 7-Methyl-4-octanone-7-$d_1$ Following Field Ionization and Following Electron Impact

| | *m/e* 87 | *m/e* 86 |
|---|---|---|
| Theoretical | | |
| Calculated assuming complete H–D randomization ($C_1$–$C_8$, and the C-7 methyl) | 56 | 44 |
| Calculated assuming H–D randomization on $C_6$–$C_8$ and the C-7 methyl only | 11 | 89 |
| Field ionization | | |
| $1 \times 10^{-11}$ sec | 100 | 0[a] |
| $2 \times 10^{-11}$ sec | 94 ± 1 | 6 ± 1 |
| $1 \times 10^{-10}$ sec | 78 ± 2 | 22 ± 2 |
| $7 \times 10^{-10}$ sec | 67 ± 6 | 33 ± 6 |
| $2 \times 10^{-6}$ sec (metastables in first field free region) | 45 ± 5 | 55 ± 5 |
| $1 \times 10^{-5}$ sec (metastables in second field free region) | 47 ± 10 | 53 ± 10 |
| Electron impact[10] | | |
| 70 eV | 65 | 35 |
| 15 eV | 61 | 39 |

[a] Obtained by extrapolation of the *m/e* 86 curve in Figure 3.

## Discussion

The "normal" FI mass spectrum of 7-methyl-4-octanone-7-$d_1$ (Figure 1b) contains no evidence of H–D randomization.[23] We attribute the *m/e* 86 peak in Figure 1b to a surface process.[20] We conclude that no H–D randomization occurs either within the neutral molecule prior to ionization or within the ionized molecule at times less than $\sim 10^{-11}$ sec following ionization. The fragmentation in the "normal" FI mass spectra of 7-methyl-4-octanone and the deuterated analog conforms to established patterns.[3-6] Fragmentations occurring at times longer than a few $\times 10^{-11}$ sec following FI are gas-phase processes within low external electric fields.[16] Following FI of 7-methyl-4-

(20) The maximum at $V_B = 8000$ V coupled with the symmetrical shape of the profile at $V_B < 8070$ V would indicate this. The evidence to date suggests that *gas-phase* rearrangement is characterized by an ion current–blade potential profile with a maximum at blade potentials higher than that required to produce the "normal" FI mass spectrum. See (a) ref 13; (b) H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, 5, 182 (1970); (c) P. Schulze and W. J. Richter, *ibid.*, 6, 131 (1971); (d) H. D. Beckey, *Z. Naturforsch. A*, 26, 1243 (1971); (e) G. Tenschert and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.*, 7, 97 (1971); (f) ref 15; (g) K. Levsen and H. D. Beckey, *ibid.*, 9, 51,63 (1972); (h) ref 16; (i) ref 17; (j) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, 95, 437 (1973); (k) ref 6.

(21) See H. D. Beckey, "Field Ionization Mass Spectrometry," Pergamon Press, New York, N. Y., 1971, pp 143–180.

(22) For a description, see L. P. Hills and J. H. Futrell, *Org. Mass Spectrom.*, 5, 1019 (1971).

(23) The "normal" FI mass spectra of deuterated hexanals contain no evidence of H–D randomization. See (a) ref 17; (b) P. Brown and C. Fenselau, *Org. Mass Spectrom.*, 7, 305 (1973).

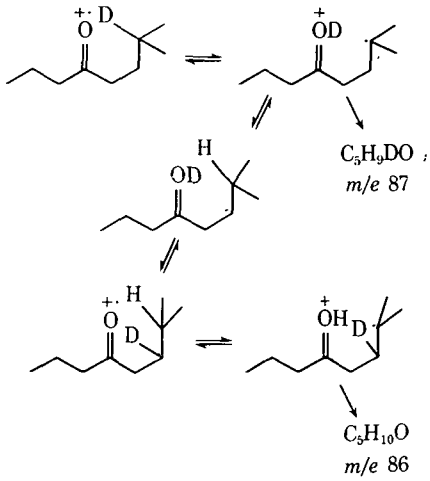
octanone the only significant fragmentation at $10^{-11}$–$10^{-9}$ sec is that leading to *m/e* 86 identified as the McLafferty rearrangement. The products of $\alpha$ cleavage, *m/e* 71 and 99, are detected at low intensities. Following FI of 7-methyl-4-octanone-7-$d_1$ two fragments, *m/e* 87 and 86, are formed at significant intensities at times of $10^{-11}$–$10^{-9}$ sec. Both fragments must be attributed to the McLafferty rearrangement. Thus either the McLafferty rearrangement is not specific or partial H–D randomization is occurring prior to fragmentation. The McLafferty rearrangement in ketones following EI is firmly established as being specific to the $\gamma$ position.[24] There are no grounds for supposing that the McLafferty rearrangement induced by FI differs in this respect.[25] We therefore attribute the *m/e* 86 fragments at times > a few $\times 10^{-11}$ sec following FI of 7-methyl-4-octanone-7-$d_1$ to partial H–D randomization prior to fragmentation. The *m/e* 86 ($C_5H_{10}O$) ion current (Figure 3) appears to drop sharply to zero at some time of the order of $10^{-11}$ sec. We interpret this behavior as an indication that there is some lower time limit within which H–D randomization does not occur prior to fragmentation. Since the time resolution of the instrument is finite (about $5 \times 10^{-11}$ sec for this experiment), the point at which the ion current drops to zero will appear at a somewhat shorter time than its true position. We place the time limit for randomization in the region of $5 \times 10^{-11}$ sec.

Following FI of 7-methyl-4-octanone-7-$d_1$ H–D randomization is able to compete successfully with the McLafferty rearrangement even within those ions sufficiently energetic to decompose at times as short as the order of $10^{-11}$ sec. If the basic tenets of the quasi-equilibrium theory[9] concerning distribution of internal energy within a molecule are valid at such short times for FI and for both high and low energy EI, then the conclusion should also hold for EI. That is to say, following both 70 and 15 eV of 7-methyl-4-octanone-7-$d_1$ H–D randomization would successfully compete with the McLafferty rearrangement at times of the order of $10^{-11}$ sec. We suggest that this is the case, and that the unusually high degree of H–D randomization manifest in the normal 70 eV EI mass spectrum can be attributed to this fact. H–D randomization in straight-chain ketones can be seen to compete successfully with fragmentation only at times of the order of $10^{-6}$ sec following FI.[6]

The relative numbers of *m/e* 87 ($C_5H_9DO$) and *m/e* 86 ($C_5H_{10}O$) formed by the McLafferty rearrangement within a few $\times 10^{-6}$ sec of FI can be estimated approximately by integrating the rate–time curves (see Experimental Section for details). We find that between $1 \times 10^{-11}$ and $5 \times 10^{-6}$ sec following FI the relative proportions of *m/e* 87 ($C_5H_9DO$) and *m/e* 86 ($C_5H_{10}O$) formed are 60 and 40%, respectively. These figures compare satisfactorily with the relative intensities observed in the EI mass spectra[10] (65 and 35% at 70 eV and 61 and 39% at 15 eV). Relative to the molecular ion intensity $I_M$, the integrated intensities of *m/e* 87 ($C_5H_9DO$) and *m/e* 86 ($C_5H_{10}O$) formed between $1 \times 10^{-11}$ and $5 \times 10^{-6}$ sec following FI are 0.03 and 0.02, respectively. The degree of fragmentation *via* the McLafferty rearrangement following FI is two orders of magnitude less than that following nominally 15 eV EI.[10] This can be attributed to a smaller amount of excitation energy imparted to the molecules by FI.

That H–D randomization following FI occurs prior to fragmentation at times as short as the order of $10^{-11}$ sec in 7-methyl-4-octanone-7-$d_1$ but not in straight-chain ketones can be accounted for on the basis of the mechanism[7] depicted in Scheme I. The mechanism is

**Scheme I**




proposed for randomization in straight-chain as well as in $\gamma$-branched ketones. The McLafferty rearrangement in aliphatic ketones at > a few $\times 10^{-11}$ sec following FI is envisaged as being stepwise[26] (*vide infra*). We suggest that the key to the success of H–D randomization in $\gamma$-branched ketones is stability of the tertiary radical a (Scheme I).[10] Consider molecular ions which ultimately decompose *via* the McLafferty rearrangement in times of a few $\times 10^{-11}$ or $10^{-10}$ sec. These will be ions of relatively high internal energies. We suggest that with a $\gamma$-branched ketone the tertiary radical intermediate is sufficiently stable to allow a degree of reverse D transfer from the oxygen to the alkyl chain (Scheme I), whereas with a straight-chain ketone the secondary radical undergoes $\beta$ cleavage before reverse transfer can occur. With a straight-chain ketone reverse transfer can only compete successfully with $\beta$ cleavage within lower energy ions decomposing at $\sim 10^{-6}$ sec.[27] Should an H transfer have occurred

(24) (a) S. Meyerson and J. D. McCollum, *Advan. Anal. Chem. Instrum.*, 2, 184 (1963); (b) H. Budzikiewicz, C. Fenselau, and C. Djerassi, *Tetrahedron*, 22, 1391 (1966).

(25) The evidence[26] to date would suggest that the reactions induced by FI at times $>10^{-11}$ sec are typically the same reactions as are induced by low energy EI (10–12 eV).

(26) The mechanism of the McLafferty rearrangement has been a contentious question. The balance of opinions in recent years would, however, seem to favor stepwise mechanisms. See (a) F. P. Boer, T. W. Shannon, and F. W. McLafferty, *J. Amer. Chem. Soc.*, 90, 7239 (1968); (b) C. Fenselau, J. L. Young, S. Meyerson, W. R. Landris, E. Selke, and L. C. Leitch, *ibid.*, 91, 6847 (1969); (c) J. S. Smith and F. W. McLafferty, *Org. Mass Spectrom.*, 5, 483 (1971). FIK measurements on hexanal suggest that in this molecule the McLafferty rearrangement following FI may be concerted but only at times of a few $\times 10^{-11}$ sec. At longer times the reaction is probably stepwise. See (d) P. J. Derrick, A. M. Falick, and A. L. Burlingame, *J. Amer. Chem. Soc.*, submitted for publication.

(27) The argument here is that low internal energies tend to favor rearrangement over direct bond cleavage. See (a) A. N. H. Yeo, R. G. Cooks, and D. H. Williams, *Chem. Commun.*, 1269 (1968); (b) A. N. H. Yeo and D. H. Williams, *J. Amer. Chem. Soc.*, 91, 3582 (1969). In comparing the $\gamma$-branched ketone to the straight-chain ketone, we are employing essentially the same argument. Replacing a secondary radical by a more stable tertiary radical can be regarded as having the same effect as lowering the internal energy. Thus the rearrangement (reverse transfer) is favored over the direct bond cleavage ($\beta$ cleavage) on introducing branching at the $\gamma$ position.

within the alkyl chain (or between the alkyl chains) prior to the reverse transfer, a step toward H–D randomization will have been effected.

The hypothesis is corroborated by the FIK data on the McLafferty rearrangements in 7-methyl-4-octanone and straight-chain ketones (Table I). In comparing the two sets of phenomenological rate constants we assume that the internal energy distributions are similar for the two molecules. We consider such an assumption to be reasonable since the two molecules are similar both in their chemical and their physical properties and since all measurements were made with the same emitter at the same field strengths and at similar temperatures. We have made FIK measurements on a large number of straight-chain and branched aliphatic ketones. The kinetics of the McLafferty rearrangement in the γ-branched 7-methyl-4-octanone differ markedly from the kinetics of the rearrangement in straight-chain ketones or branched ketones in which there are secondary hydrogens at the γ position. The phenomenological rate constants for the McLafferty rearrangement in 7-methyl-4-octanone at times of $10^{-11}$–$10^{-10}$ sec following FI are approximately an order of magnitude less than the corresponding phenomenological rate constants for 2-octanone. At times of $\sim 10^{-6}$ sec the situation is reversed with the rate constants for 7-methyl-4-octanone being approximately an order of magnitude greater. The total fraction of ions undergoing the McLafferty rearrangement following FI is, however, approximately the same ($\sim 5 \times 10^{-2}$) for both compounds[28] (based on integration of the rate–time curves). The situation can be summarized by saying that although the McLafferty rearrangement is of similar importance in both compounds, it is considerably *slower* in the branched compound.[29] The gross differences between the kinetics of the two McLafferty rearrangements cannot be reasonably explained on the bases of mechanisms representing the reactions as concerted processes. We will thus consider that the reactions occur in a stepwise fashion (at least at times > a few $\times 10^{-11}$ sec). It is unlikely that the differences between kinetics arise in the hydrogen transfers, since it is unreasonable to propose that the transfer of a labile tertiary hydrogen be considerably slower than the transfer of a secondary hydrogen. If the differences between the kinetics do not arise in the hydrogen transfers, they presumably arise in the subsequent β-cleavage steps. The β cleavage in the γ-branched ketone must be slower than that in the straight-chain ketone, which is consistent with the tertiary radical a (Scheme I) being more stable than the corresponding secondary radical in the straight-chain ketone. A slower β cleavage in 7-methyl-4-octanone supports our proposals as to the mechanism of H–D randomization.

Further support for the hypothesis can be marshalled from FIK measurements[30] on 3-methyl-2-heptanone-*1,1,1,3-d*$_4$. It is found that no H–D randomization occurs prior to the McLafferty rearrangement at times less than $1 \times 10^{-9}$ sec following FI. H–D randomization in this compound following FI bears the same characteristics as that in a straight-chain ketone.[6] This is as would be predicted from the proposed hypothesis, since γ-H transfer in 3-methyl-2-heptanone-*1,1,1,3-d*$_4$ gives rise to a secondary radical.[31] We would point out that the failure of α branching and α labeling to enhance H–D randomization in the same manner as γ branching and γ labeling weighs against mechanisms[8] of the type which involve simultaneous migration of an H and a D effecting a direct exchange.[32]

It remains only to comment on the dependence of the H–D randomization process in 7-methyl-4-octanone upon internal energy *E*. It is clear (Table II) that the degree of H–D randomization prior to the McLafferty rearrangement following FI increases with time. Making the reasonable assumption that reaction time is related in a direct manner to internal energy *E*, it follows that the success of the H–D randomization process in competing with the McLafferty rearrangement increases as internal energy *E* decreases. Following EI the degree of H–D randomization in 7-methyl-4-octanone-*7-d*$_1$ prior to the McLafferty rearrangement does increase to a small extent on lowering the electron energy from 70 to 15 eV (Table II).[33] It is interesting that the intensities of *m/e* 86 ($C_5H_{10}O$) formed in the first and second field-free regions by the McLafferty rearrangement following FI of 7-methyl-4-octanone-*7-d*$_1$ are higher than the calculated intensities for complete H–D randomization (Table II). This probably indicates that complete H–D randomization has not been reached. Partial randomization of just the hydrogens attached to $C_6$–$C_8$ and the C-7 methyl, for example, could produce this result. A kinetic isotope effect may also be present, but replacement of γ-H by D does not produce a kinetic isotope effect on the McLafferty rearrangement in aliphatic ketones following 70 eV ionization.[34]

Our discussion has been directly concerned only with reactions induced by FI at times of $10^{-11}$–$10^{-5}$ sec. We would suggest, however, that the proposals as to mechanisms of H–D randomization and of the McLafferty rearrangement are equally valid within the same time frame following low energy EI and quite possibly high energy EI.

(28) The degrees of fragmentation *via* the McLafferty rearrangement in the EI mass spectra of 7-methyl-4-octanone and 2-octanone are similar.[7,10]

(29) Some figures based on the aforementioned integration of rate–time curves exemplify this. Only approximately 25% of the total fragmentation of 7-methyl-4-octanone *via* the McLafferty rearrangement following FI occurs in times less than $1 \times 10^{-9}$ sec, whereas 75% occurs at times longer than $1 \times 10^{-9}$ sec; the corresponding figures for 2-octanone are 60 and 40%, respectively.

(30) P. J. Derrick, A. M. Falick, and A. L. Burlingame, unpublished data.

(31) δ-Branched and δ-D labeled ketones exhibit enhanced H–D randomization prior to the McLafferty rearrangement but not to the same degree as the γ branched and γ labeled.[10] We would suggest that in this case a process of the type represented in Scheme I is still operative but involving δ transfer *via* a seven-membered cyclic transition state facilitated by the labile D. Hydrogen transfers in carbonyl compounds *via* seven-membered cyclic transition states have been suggested previously. See (a) ref 7; (b) R. J. Liedtke and C. Djerassi, *J. Amer. Chem. Soc.*, **91**, 6814 (1969); (c) M. Kraft and G. Spiteller, *Org. Mass Spectrom.*, **2**, 541 (1969); (d) A. G. Harrison, *ibid.*, **3**, 549 (1970).

(32) FIK and EI measurements on 2-methylpropane suggest that such exchange-type processes do operate with this compound but are favored by high excitation energies (P. J. Derrick, A. M. Falick, and A. L. Burlingame, in preparation for *J. Chem. Soc., Perkin Trans. 2*,) whereas H–D randomization even in γ-branched ketones is favored by low energies.

(33) The extent of H–D randomization prior to α cleavage of 7-methyl-4-octanone-*7-d*$_1$ following EI appears to be independent of electron energy.[10] The reason for this is not obvious from the present experimental data. The effect that varying the electron energy has on the competition between the α cleavages and the McLafferty rearrangement may be relevant to the problem.

(34) J. K. Macleod and C. Djerassi, *J. Amer. Chem. Soc.*, **89**, 5812 (1967).

## Experimental Section

FIK measurements were made with a modified double focusing Du Pont (CEC) 21-110B mass spectrometer.[16] Uncoated Schick stainless steel razor blades were used as emitters in the standard Du Pont FI/EI ion source. Curves of ion current *vs.* blade potential were measured on an XY recorder. The blade potential was scanned from 8 to 10 kV in steps of 10 V so that the curves are quasi-continuous.

The synthesis of the deuterated compound has been described.[10]

Total fragment ion currents formed between $1 \times 10^{-11}$ and $5 \times 10^{-6}$ sec following FI were estimated as follows. The logarithm of the rate constant $\bar{k}(t)$ calculated from eq 1 for the particular reaction of interest was plotted against log $t$. The resulting graph had a series of points in the range $10^{-11}$–$10^{-9}$ sec and two additional points at $1 \times 10^{-5}$ and $2 \times 10^{-6}$ sec (from the metastables). A curve was drawn through all these points *interpolating between* $10^{-9}$ *and* $2 \times 10^{-6}$ *sec*. With both *m/e* 86 ($C_5H_{10}O$) and *m/e* 87 ($C_5H_9DO$) from 7-methyl-4-octanone-7-$d_1$ these curves of log $\bar{k}(t)$ *vs.* log $t$ could be approximated by straight lines. The area under the curves could then be estimated by analytical integration. The area under such a curve represents the total current due to a particular fragment ion (expressed as a fraction of the molecular ion current) formed between the time limits of integration.

**Acknowledgment.** The financial support of the National Aeronautics and Space Administration (Grant NGL 05-003-003) and the National Science Foundation (Grant NSF GP-38389X) for this work is gratefully acknowledged.

[redacted]

474-44059

*Reprinted from*

OMS **Organic Mass Spectrometry**
AN INTERNATIONAL JOURNAL

# ORGANIC MASS SPECTROMETRY

**EDITOR-IN-CHIEF**
PROF. ALLAN MACCOLL

**Assistant Editor:** DR. A. G. LOUDON

*University College, 20 Gordon Street, London WC1H0AJ*

**EDITORIAL ADVISORY BOARD**

Dr. J. A. Ballantine, *University College of Swansea;* Prof. H. D. Beckey, *University of Bonn;* Prof. J. H. Beynon FRS, *ICI, Manchester;* Prof. K. Biemann, *Massachusetts Institute of Technology;* Prof. H. Budzikiewicz, *University of Cologne;* Dr. N. R. Daly, *A.W.R.E., Aldermaston;* Prof. C. Djerassi, *Stanford University;* Dr. H. W. Fehlhaber, *Farbwerke Hoechst AG, Frankfurt;* Prof. M. Fetizon, *University of Paris;* Dr. V. Hanus, *Czechoslovak Academy of Sciences, Prague;* Dr. J. L. Holmes, *University of Ottawa;* Dr. W. Kelly, *Unilever Research Laboratory, Bedford;* Prof. F. W. McLafferty, *Cornell University;* Dr. A. Mandelbaum, *Technion, Haifa;* Dr. S. Meyerson, *Standard Oil Company (Indiana), Naperville, Illinois;* Dr. B. J. Millard, *School of Pharmacy, London;* Prof. K. Nakanishi, *Columbia University;* Dr. R. I. Reed, *University of Glasgow;* Dr. K. D. Schuy, *RPI-Hilger & Watts, Germany;* Dr. D. H. Williams, *Cambridge University.*



*Publishing Offices:*
Spectrum House, Alderton Crescent, London NW4 3XX

*Production Editor:* Celia Gould

One volume *per annum.* Published Monthly.
1973 Subscription rates (incl. postage): £49·00 $120.00 DM 380.00

Back issues Vols. 1–6 available.



© Heyden & Son Ltd, 1973

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Heyden & Son Ltd.



**REGIONAL EDITORS**

**North America:** Prof. Robert H. Shapiro, *Department of Chemistry, University of Colorado, Boulder, Colorado* 80302

**Associate Editor:** Dr. S. P. Markey, *Department of Pediatrics, University of Colorado Medical Center, Denver, Colorado* 80220

**France:** Prof. A. Cornu, *Commissariat a l'Énergie Atomique, Centre d'Études Nucléaires, Cedex No.* 85, *Grenoble* 38

**Germany:** Prof. H.-Fr. Grützmacher, *Institut für Organische Chemie, der Universität,* 2 *Hamburg* 13, *Papendamm* 6

**Netherlands:** Prof. G. Dijkstra, *Analytisch Chemisch Laboratorium der Rijksuniversiteit Utrecht, Croesestraat* 77a, *Utrecht*

**Assistant Editor:** Dr. J. J. de Ridder, *N. V. Organon, Kloosterstraat* 6, *Oss*

**USSR:** Prof. N. K. Kochetkov, *N.D. Zelinsky Institute of Organic Chemistry, Academy of Sciences of the USSR, Moscow* B 334, *Leninsky prospect* 47

**Australasia:** Prof. J. S. Shannon, *Chemistry School, University of New South Wales, P.O. Box* 1, *Kensington, N.S.W.* 2033, *Australia*

**Japan:** Prof. Akira Tatematsu, *Faculty of Pharmacy, Meijo University, Showa-ku, Nagoya*



Heyden & Son Limited, Spectrum House, Alderton Crescent, London NW4 3XX
(Tel: 01-202 5333; Telex: 28303)
Heyden & Son GmbH, 4440 Rheine/Westfalen, Postfach 529, Germany
(Tel: (02531)8111; Telex: 893870)

Organic Mass Spectrometry, 1974, Vol. 8, pp. 279 to 287. Heyden & Son Limited. Printed in Northern Ireland

# APPARENT CHARGE LOCALIZATION EFFECTS IN THE McLAFFERTY REARRANGEMENT OF ACETATE ESTERS: FIELD IONIZATION KINETIC MEASUREMENTS

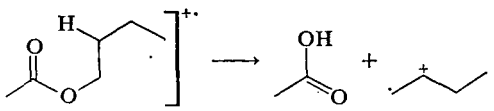
GORDON WOOD,* A. M. FALICK† and A. L. BURLINGAME
Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

(*Received* 11 *October* 1972; *accepted* (*revised*) 2 *July* 1973)

**Abstract**—The relative rates of the McLafferty rearrangement in *n*-butyl acetate and several related molecules have been examined over the time range $10^{-11}$ to $10^{-5}$ s by means of a field ionization mass spectrometric technique. It was found that the introduction of a substituent group of low ionization potential (phenyl, dimethylamino) at sites remote from the ester oxygens reduces the rate of the McLafferty rearrangement and that the rate becomes lower as the substituent is moved further away from the ester group. The results suggest that for the cases examined the charge on the molecular ion may be localized on the functional group of lowest ionization potential.

## INTRODUCTION

PREVIOUS work from this laboratory utilizing field ionization mass spectrometry (FIMS) has shown that *n*-butyl acetate undergoes a rapid, site-specific, gas phase McLafferty type rearrangement[1] which may be represented as follows:



(1)

The electron-impact (EI) mass spectra of many compounds containing functional groups related to esters show abundant 6-center rearrangement ions corresponding to the McLafferty rearrangement. Most of these molecules, however, give small or undetectable rearrangement peaks in normal field ionization (FI) mass spectra‡ taken with a double focusing instrument. Since only fragment ions formed at times less than about $10^{-11}$ s after ionization (and ions formed at the emitter surface) are normally detected in such spectra,[1,2] this implies that the McLafferty rearrangement in aliphatic esters proceeds at an unusually rapid rate.[3]

In an attempt to provide some further insight into the nature of this rearrangement we have employed the field ionization kinetic (FIK) method.[2,4,5,6] This technique was used because it can afford much more detailed kinetic information about mass spectrometric fragmentation processes than conventional EI or FI spectra (e.g. see Refs. 7 to 9). It provides the capability of measuring the rates of unimolecular decomposition of primary ions at times of $10^{-11}$ to $10^{-5}$ s after ionization. The fact that the field ionization process imparts comparatively little excitation energy to the molecular ions tends to increase the relative contribution of low frequency factor, low activation energy processes (e.g. rearrangements) to the total fragmentation.[7,10,11]

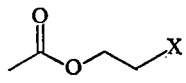
* Present address: Department of Chemistry, University of Windsor, Windsor, Ontario, Canada.
† Author to whom correspondence should be addressed.

‡ A number of amides, ethers, ester vinylogs and nitrogen and sulfur compounds closely related to esters were examined in this laboratory. None was found to have a significant peak in its FI spectrum corresponding to a McLafferty type rearrangement.

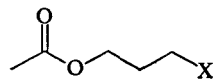
FIK would therefore appear to be an excellent method for study of the effects of structural changes at a site remote from the carbonyl group in Reaction (1).

## RESULTS

The compounds chosen for study all contain the hydrogen atom at C-2 which had been implicated in the butyl acetate rearrangement[1] and in addition, a functional group (either phenyl or dimethylamino) further along the alcohol chain.



(I) X = $C_2H_5$
(II) X = $C_6H_5$
(III) X = $N(CH_3)_2$



(IV) X = $C_6H_5$
(V) X = $N(CH_3)_2$

Figure 1 shows the phenomenological rate constants $\bar{k}(t)$ plotted as a function of parent ion lifetime for the loss of a neutral fragment of mass 60 from compounds I to

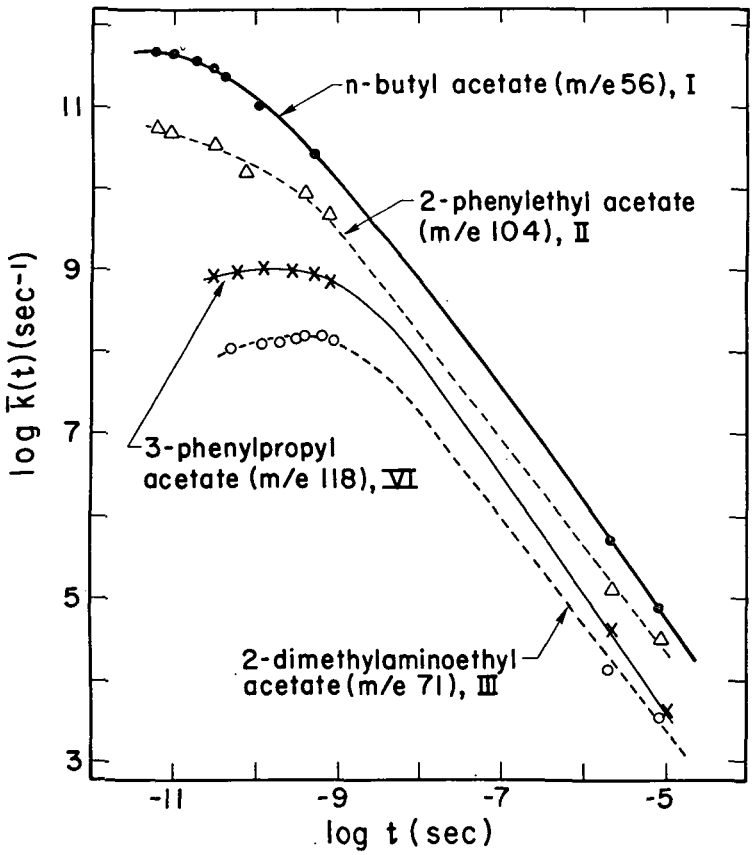
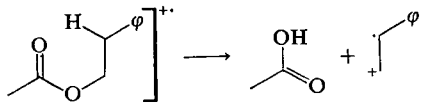



FIG. 1. Log-log plot of phenomenological rate constant $\bar{k}(t)$ (plotted as a function of parent ion lifetime) for the loss of a neutral fragment of *m/e* 60 from compounds I to IV following field ionization.

IV following field ionization. The rate of this reaction for compound V was so slow that no product ion (*m/e* 85) resulting from gas phase decomposition was detectable in any portion of the time range covered in Fig. 1. The data clearly show that the rate of the rearrangement for compounds I to V is in the order I > II > IV > III > V over the entire time range shown in the figure. The maximum rates appear to fall in the same order, although in the case of compounds I and II these maxima lie at times shorter than $\sim 10^{-11}$ s and could not be directly measured.

It is interesting to note that in the case of II and III the rate constants are reduced despite an apparent opportunity for charge delocalization in the product ion. For example, the charged fragment from II is formally equivalent to ionized styrene:




(2)

Compound IV no longer contains this apparent advantage and here one should be looking at a more clear-cut example of introduction of a group remote from the presumed site of ionization. Here the maximum rate is much lower than for *n*-butyl acetate and occurs at a time at least 2 orders of magnitude longer.

Figure 2 shows the low energy ($\sim$ 18 eV) EI spectrum of *n*-butyl acetate (see Ref. 1b for the FI spectrum). Figures 3 to 6 contain the low energy EI spectra and the normal FI spectra of compounds II to IV. The FI spectra must be interpreted with

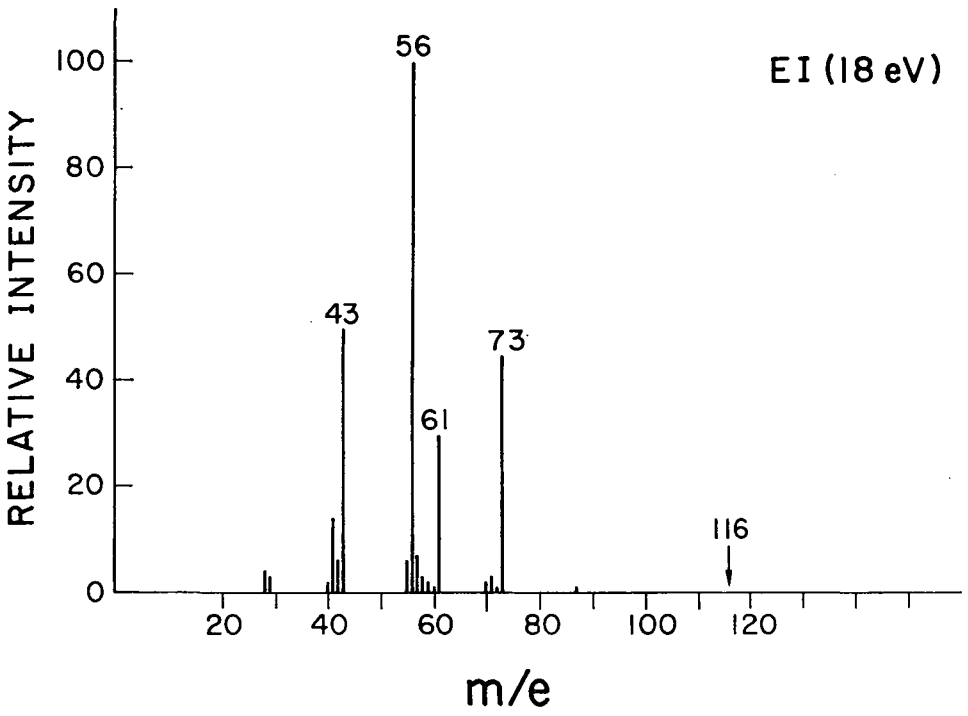



FIG. 2. Low energy ($\sim$ 18 eV) EI mass spectrum of *n*-butyl acetate (I).

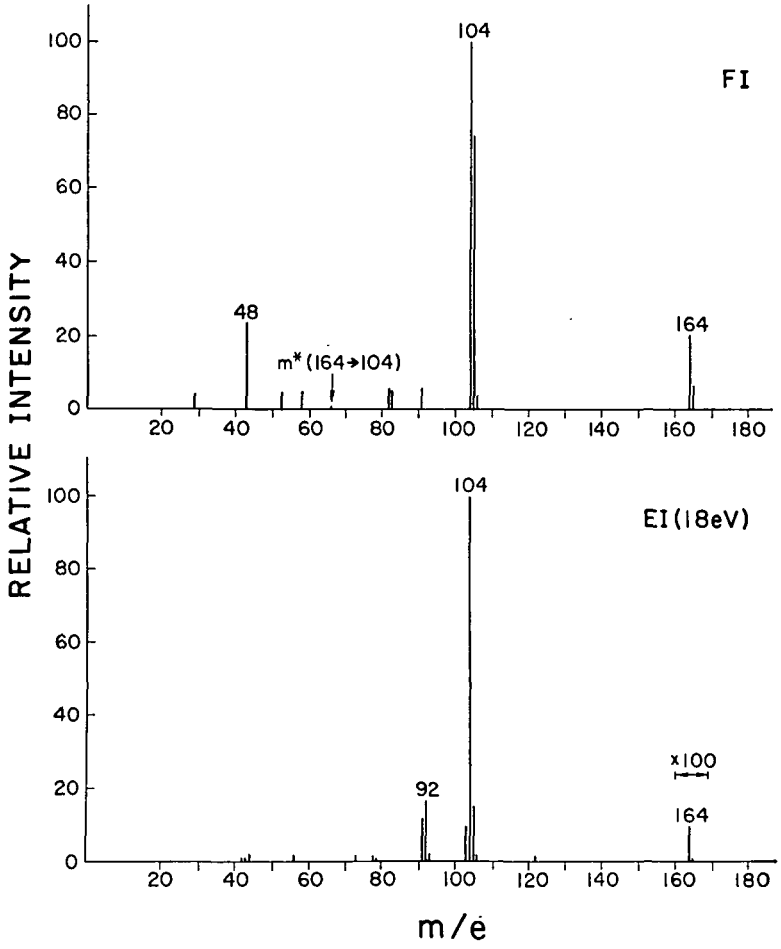



FIG. 3. Low energy (~ 18 eV) EI and normal FI mass spectra of 2-phenylethyl acetate (II).

caution, because they represent fragments produced via surface and high field processes[5,12] as well as those produced in the gas phase at times up to $\sim 10^{-11}$ s following ionization.[2] For example, the small *m/e* 85 peak observed in the FI mass spectrum of V is very probably due to a surface and/or high field process since no evidence of gas phase reaction was observed at times slightly longer than $10^{-11}$ s. Thus, peaks in these FI spectra must be regarded as upper limits to the amount of gas phase reaction which occurs during the first $10^{-11}$ s of the parent ion lifetime.

## DISCUSSION

The differences in the rates of the McLafferty rearrangement* in the molecules studied are most readily interpreted in terms of the preferential location of charge on

* It should be noted that the particular rearrangement examined here differs slightly from the usual McLafferty rearrangement which results in a charged acid fragment. In the present case, a sort of 'inverse' McLafferty process (in which the charge is retained on the olefin fragment) is operative. Recent evidence (P. J. Derrick, A. M. Falick and A. L. Burlingame, to be published) suggests that these are two distinct reactions and that the latter reaction has the lower activation energy.

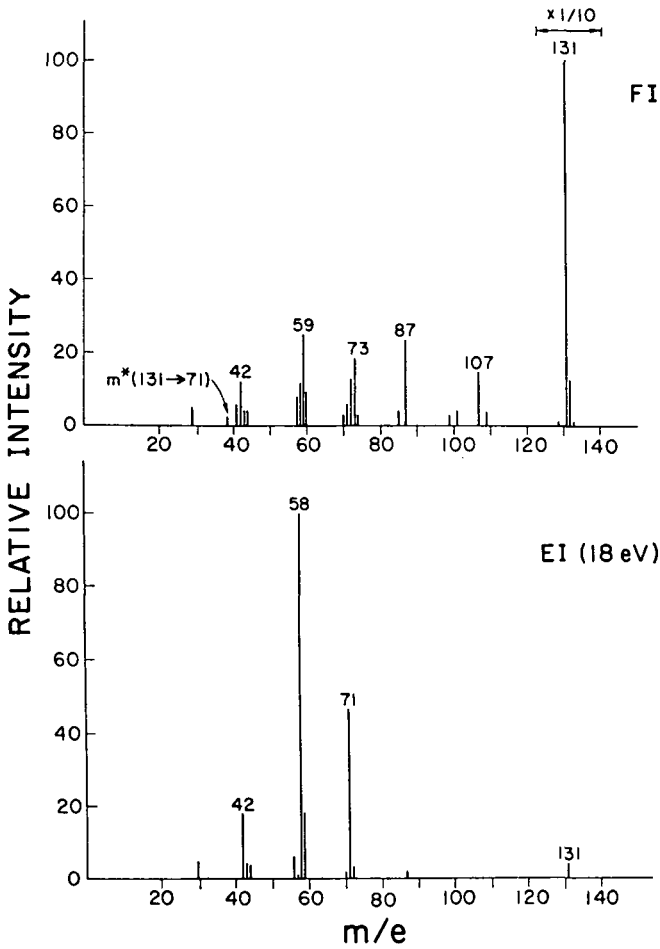



FIG. 4. Low energy (~ 18 eV) EI and normal FI mass spectra of 2-dimethyl-aminoethyl acetate (III).

the group of lower ionization potential. Typical values of the first ionization potential for aliphatic esters are about 10·0 eV (Ref. 13), whereas the ionization potential of benzene is 9·25 eV (Ref. 13) and a typical value for aliphatic amines is 8·7 eV (Ref. 13). Thus our results are consistent with the expectation that a charge localization effect will be present in compounds II and IV, and more pronounced in compounds III and V.

As Howe and Williams[14,15] have pointed out, however, interpretation in these terms is not necessarily in conflict with a quasi-equilibrium theory (QET)[16,17] approach. Ions formed by FI have relatively small internal energies.[18] If it is assumed that the accessible electronic states are in equilibrium, then those molecular ions occupying the lowest lying electronic state will also have the highest vibrational energies and therefore this electronic state will have the greatest degeneracy. Decompositions are most likely to occur from the lower electronic states on both statistical and energetic grounds. The lowest electronic state of a molecular ion corresponds to the removal of one

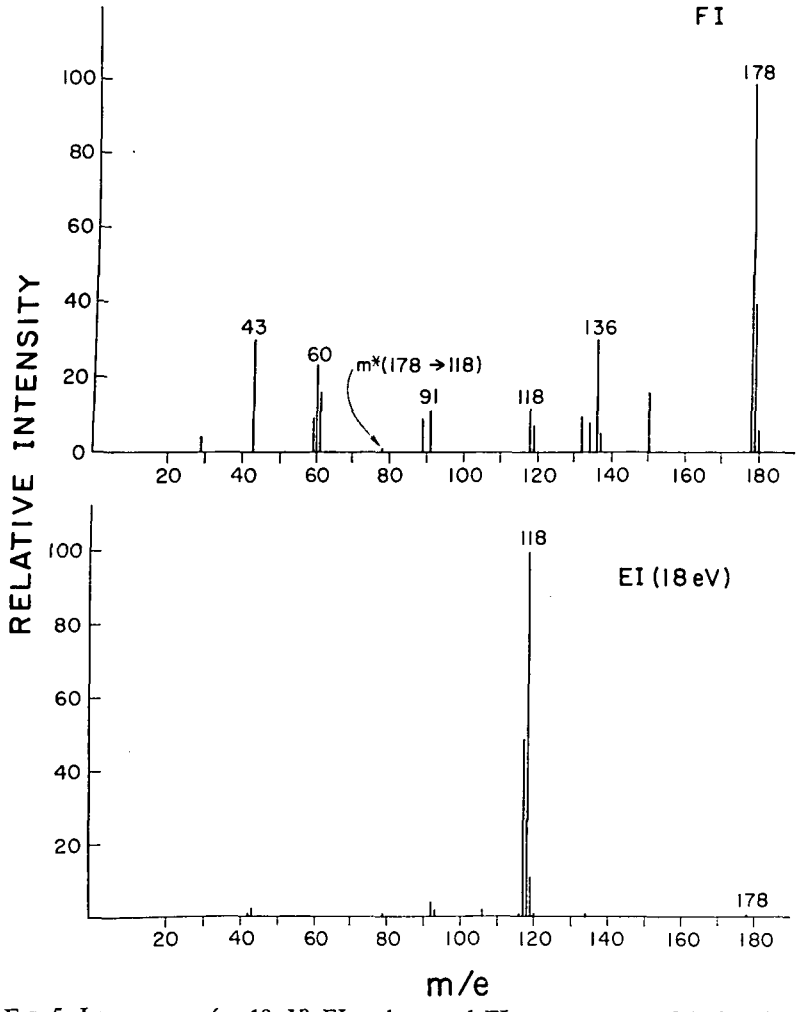



FIG. 5. Low energy (~ 18 eV) EI and normal FI mass spectra of 3-phenylpropyl acetate (IV).

electron from the highest occupied molecular orbital. It is precisely this state which would be expected to have the greatest amount of charge concentrated at the part of the molecule having the lowest ionization potential.

For the dimethylamino substituted series (I, III, V) the relative intensities of the rearrangement ions from Reaction (1) in the EI spectra (Table 1) suggest that charge localization or its equivalent is operative. Thus both 70 eV and 18 eV spectra show an apparent small reduction in the rearrangement ion upon introduction of the dimethylamino group at C-2 (III *vs.* I) and a further reduction of about one order of magnitude when the dimethylamino group is moved to C-3 (V *vs.* III). By comparison, the most energetic ions detected in the FIK measurements (those decomposing at $\sim 10^{-11}$ s after ionization) show a difference in decomposition rate of about 3 orders of magnitude between III and I. In the case of the phenyl substituted series (IV *vs.* II *vs.* I), no charge localization effect can be discerned from the EI spectra, whereas the FIK data show successive changes in decomposition rate of one to two orders of magnitude

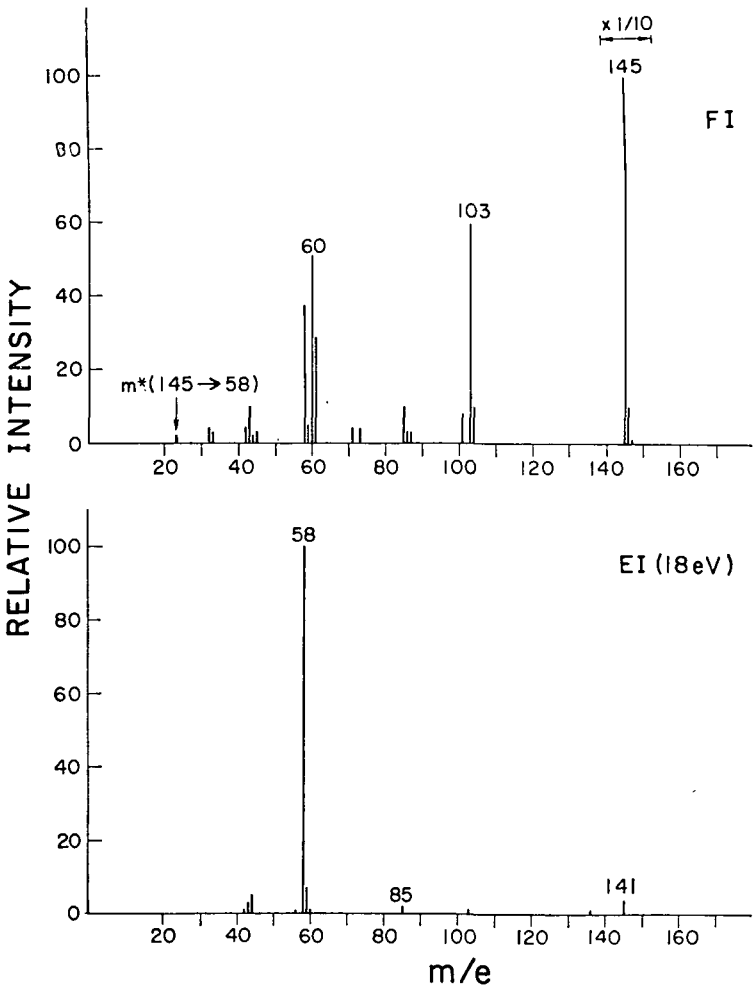



FIG. 6. Low energy (~ 18 eV) EI and normal FI mass spectra of 3-dimethylamino-propyl acetate (V).

The greatly reduced ability to distinguish the charge localization effect following EI ionization as compared to FI is very likely the result of the increased internal energy imparted to the ions in this case. The fact that the effect is substantial in the spectra of the dimethylamino compounds lends support to this hypothesis in view of

TABLE 1. INTENSITIES OF FRAGMENT IONS PRODUCED IN REACTION (1) (AS A PERCENTAGE OF TOTAL IONIZATION) FOLLOWING 70 eV AND 18 eV EI IONIZATION. THE COMPOUNDS ARE LISTED IN DECREASING ORDER OF RATE OF REACTION (1) AS MEASURED BY FIK

| Compound | *m/e* | 70 eV | 18 eV |
|---|---|---|---|
| (I) | 56 | 20 | 36 |
| (II) | 104 | 5 | 59 |
| (IV) | 118 | 7 | 58 |
| (III) | 71 | 15 | 24 |
| (V) | 85 | <1 | 1 |

the respective ionization potentials for these groups. An additional factor which may contribute to the lack of a clear-cut trend in the series I, II, IV is the expected strong activation of the rearranging hydrogen by the adjacent benzene ring in compound II. The larger amount of energy available following EI ionization, particularly by 70 eV electrons, as compared to FI[18] causes a variety of higher activation energy processes (e.g. direct cleavage reactions) to compete successfully with Reaction (1), thereby depleting the supply of molecular ions at an unknown rate. Furthermore, the additional available internal energy undoubtedly enhances secondary fragmentation of the McLafferty product ion to an unknown but almost certainly significant degree. Unfortunately, the kinetic details of processes which occur at times less than $\sim 10^{-6}$ s after ionization cannot be examined by means of conventional EI techniques. The intensities of particular peaks in EI mass spectra represent only the integrated result after $\sim 10^{-6}$ s of all of the competing and consecutive reactions leading to and from the product of interest.

The FIK measurements, on the other hand, begin with a lower energy ionization process which suppresses competing reactions of the molecular ion and subsequent reactions of the product ion. The extent of depletion of the molecular ion can be calculated since only one or two fragmentations occur to a significant extent and the rates of these decompositions can be measured as functions of time.

A good deal of study has gone into the question of charge localization in EI mass spectrometry.[19] The present work suggests that FIK measurements can be a valuable complementary tool in clarifying mechanistic questions of this kind. In particular, it has been shown that when the fragmentations of field ionized molecules are examined in such a way that surface and high field effects are eliminated, the results are consistent with reasonable chemical expectations.

## EXPERIMENTAL

The experiments were performed with a modified[2] Du Pont (CEC) 21-110B double focusing mass spectrometer equipped with a combination FI/EI ion source. Uncoated Schick stainless steel razor blades were used as FI emitters. All spectra were produced at 80°C. Decompositions occurring in the $10^{-11}$ to $10^{-8}$ s region ('fast metastables') were detected by means of a defocusing technique essentially similar to the 'metastable defocusing technique',[20,21] in which the accelerating potential was raised to permit ions formed at varying distances from the blade surface to be detected.[1b,2] The resulting curves of ion current at a particular mass against accelerating potential were transformed to yield rates of dissociations as a function of parent ion lifetime.[2] Dissociations occurring in the first field free region (lifetimes $\sim 10^{-6}$ s) were detected by the metastable defocusing technique.[20,21] Ions formed by dissociations in the second field free region ($\sim 10^{-5}$ s) appeared as normal metastable peaks in the FI spectrum.

The low energy EI spectra were obtained on a Varian MAT 311 mass spectrometer coupled to an XDS Sigma 2 computer.[22]

The compounds used in this study were obtained commercially with the exception of 3-dimethylaminopropyl acetate. *n*-Butyl acetate (I) and 2-dimethylaminoethyl acetate (III) were obtained from Eastman. 2-Phenylethyl acetate (II) was obtained from K & K. 3-Phenylpropyl acetate (IV) was obtained from Aldrich.

Purity of all samples was checked by gas liquid chromatography (g.l.c.) and EI mass spectrometry.

3-Dimethylaminopropyl acetate (V) was prepared from acetic anhydride and 3-dimethylamino propanol by heating on a steam bath. The product was isolated by g.l.c..

*Acknowledgements*—The authors are indebted to Dr Robert Cox for obtaining the low energy EI spectra, and to the National Aeronautics and Space Administration for financial support (Grant NGL 05-003-003). One of us (GW) wishes to thank the University of Windsor for a sabbatical leave during which this work was initiated.

## REFERENCES


1. (a) A. L. Burlingame, A. M. Falick, G. W. Wood, P. Schultze and W. J. Richter, American Society for Mass Spectrometry, 18th Annual Conference on Mass Spectrometry and Allied Topics, San Francisco, California, June 1970, p. 416; (b) P. Schultze and W. J. Richter, *Int. J. Mass Spectrom. Ion Phys.* **6,** 131 (1971).
2. A. M. Falick, P. J. Derrick and A. L. Burlingame, *Int. J. Mass Spectrom. Ion Phys.* in press.
3. K. Levsen and H. D. Beckey, *Int. J. Mass Spectrom. Ion Phys.* **7,** 341 (1971).
4. H. D. Beckey, H. Hey, K. Levsen and G. Tenschert, *Int. J. Mass Spectrom. Ion. Phys.* **2,** 101 (1969).
5. H. D. Beckey, *Field Ionization Mass Spectrometry*, Pergamon Press, New York, 1971.
6. P. J. Derrick and A. J. B. Robertson, *Proc. Roy. Soc.* (*A*) **324,** 491 (1971).
7. P. J. Derrick, A. M. Falick and A. L. Burlingame, *J. Amer. Chem. Soc.* **94,** 6794 (1972).
8. P. J. Derrick, A. M. Falick and A. L. Burlingame, *J. Amer. Chem. Soc.* **95,** 437 (1973).
9. P. J. Derrick, A. M. Falick, S. Lewis and A. L. Burlingame, *Org. Mass Spectrom.* **7,** 887 (1973).
10. A. N. H. Yeo, R. G. Cooks and D. H. Williams, *Chem. Commun.* 1269 (1968).
11. A. N. H. Yeo and D. H. Williams, *J. Amer. Chem. Soc.* **91,** 3582 (1969).
12. H. D. Beckey and F. W. Röllgen, *Naturwissenschaften* **58,** 23 (1971).
13. J. L. Franklin, J. G. Dillard, H. M. Rosenstock, J. T. Herron, K. Draxl and F. H. Field, *Ionization Potentials. Appearance Potentials, and Heats of Formation of Gaseous Positive Ions*, Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. No. 26, 1969.
14. I. Howe and D. H. Williams, *J. Amer. Chem. Soc.* **90,** 5461 (1968).
15. I. Howe, in D. H. Williams (Sr. Reporter), *Mass Spectrometry*, Vol. 1, The Chemical Society, London, p. 80.
16. H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig and H. Eyring, *Proc. Natl. Acad. Sci. U.S.* **38,** 667 (1952).
17. H. M. Rosenstock, in E. Kendrick (Ed), *Advances in Mass Spectrometry*, Vol. 4, The Institute of Petroleum, London, 1968, p. 523.
18. H. D. Beckey, in K. Ogata and T. Hayakawa (Eds), *Recent Developments in Mass Spectroscopy*, University Park Press, Baltimore, 1970, p. 1154.
19. For a recent paper dealing specifically with the McLafferty rearrangement see: J. R. Dias, Y. M. Sheikh and C. Djerassi, *J. Amer. Chem. Soc.* **94,** 473 (1972).
20. M. Barber and R. M. Elliott, American Society for Mass Spectrometry, 12th Annual Conference on Mass Spectrometry and Allied Topics, Montreal, Canada, June 1964.
21. P. Schultze and A. L. Burlingame, *J. Chem. Phys.* **49,** 4870 (1968).
22. D. H. Smith, R. W. Olsen, F. C. Walls and A. L. Burlingame, *Anal. Chem.* **43,** 1796 (1971).

# INFORMATION FOR CONTRIBUTORS

## General

*Organic Mass Spectrometry* is devoted to the rapid publication of those papers exclusively dealing with the application of mass spectrometry to problems in organic chemistry. Intended papers should clearly add original scientific value to the field. Contributions should not have been published elsewhere and authors must agree not to communicate the same material for publication in any other journal or book. Authors are invited to submit their papers for publication to the Regional Editor where appropriate, or in all other cases to the Editor-in-Chief. From time to time review articles will be accepted, but prior consultation with one of the Editors is strongly advised. Papers may be in English, French or German. Safe receipt of contributions for publication will be acknowledged by the receiving Editor.

## Manuscripts

Manuscripts should be typed in double line spacing, and margins of at least 1½ inches must be left at the top, bottom and left-hand side of each page. The first page should also bear the name and full address of one author to whom the page proofs will be despatched in due course. All manuscripts should be submitted in triplicate, and in addition, a set of copies or prints of the diagrams must be attached to each copy of the manuscript. This will enable papers to be sent to two referees simultaneously and so assist rapid publication.

## Abstract

Every paper for publication must be preceded by an abstract of not more than 250 words setting out the essential contents. For French and German papers an English version of the abstract should also be provided.

## Nomenclature

Authors should conform to nomenclature, symbols and abbreviations adopted by the Commission on the Nomenclature of Organic Chemistry of the International Union of Pure and Applied Chemistry (IUPAC).

## Illustrations

Illustrations on tracing cloth, tracing paper or white card should be supplied separately from the text and must bear the names of the authors, the title and the figure number. The mass numbers for principal peaks should be indicated. Such illustrations should be carefully drawn at approximately twice the size they are required to be reproduced in the actual journal. This reduction is necessary to ensure high definition in print. Lettering must be large enough to withstand this 50% reduction and remain clearly legible. Printing area of a page is 8″ × 5″. Illustrations should be provided with descriptive legends which should be listed on a separate sheet.

## Tables

Tables should be kept to the minimum. The same data should not be reproduced in both tabular and diagrammatic form. Essential tables should be provided with legends making them intelligible without reference to the text.

## Experimental

New compounds should be indicated by underlining the name (for italics) at its first mention (excluding headings) in the Experimental Section only, and by giving analytical results in the form: (Found: C, 63·1; H, 5·4. $C_{13}H_{13}NO_4$ requires C, 63·2; H, 5·3%). Analytical results for compounds which have previously been adequately described in the literature should be given in the form: (Found: C, 62·95; H, 5·4. Calc. for $C_{13}H_{13}NO_4$: C, 63·2; H, 5·3%). The punctuation should be noted.

## References

References should only be given for available published work, and will be cited in the text by superior numbers and listed at the end of the article. They should be in the following form:

1. R. L. Baird and A. Aboderin, *J. Amer. Chem. Soc.* **86**, 252 (1964).
2. H. Budzikiewicz, C. Djerassi and D. H. Williams, *Interpretation of Mass Spectra of Organic Compounds*, Holden-Day, San Francisco, 1964.

Journal titles must be abbreviated according to the system adopted by *Chemical Abstracts*. (Supply full titles otherwise.)

## Footnotes

Footnotes, as distinct from literature references, should be indicated by the following symbols:

* † ‡ § ‖ ¶

used in this sequence and commencing anew on each page; they should not be included in the numbered reference system.

## Acknowledgements

These are placed at the end of the text preceding the references and should be brief.

## Proofs

Page proofs, for the correction of printers' errors only will be despatched to the author indicated on the first page of the manuscript. Any further alteration at this proof stage may have to be charged to the author and may seriously delay publication. Fifty reprints of each paper will be supplied free of charge. Additional copies may be ordered if required on a form supplied with the proofs.

## Short Communications

Short communications will appear as a distinct section of the journal following the ordinary papers. Reprints will be available at the normal rates.

In style and presentation short communications should conform to that laid down for manuscripts, but length should be restricted to a maximum of two journal pages, i.e. approximately four pages of typed manuscript inclusive. Graphical spectral diagrams will not normally be published and if necessary details should be tabulated. Reaction schemes must be kept as simple as possible and drawn suitable for direct reproduction. The scale factor should be 1½:1 i.e. not more than 7½″ (17·5 cm) wide for reduction to normal page width. Journal style should be taken into account. Editors will be responsible for refereeing and to enable as rapid publication as possible authors proofs will not be awaited before passing for press unless specifically requested, although they will be sent as a matter of course and corrections incorporated if notified in time.

*Short communications should still be preceded by an abstract as for normal papers.*

HEYDEN & SON LTD

journal of dynamic systems, measurement, and control

forum

N77 — 84188

A continuing feature in which noted individuals in various areas are solicited to report on what is and should be going on in their fields. Response is welcomed and will appear in the Discussion section of future issues.

# Systems Analysis as Technology Transfer

IDA R. HOOS[1]

## Semantics of Systems Analysis

NOT since M. Jourdain, the untutored tradesman in Molière's play, *Le Bourgeois Gentilhommc,* made the happy discovery that prose was what he had been talking all his life has there appeared so ubiquitous and so handy a talent as "systems capability." In its common and current usage, this term denotes the ability to study, design, devise, revise, and/or analyze systems. And here we encounter the first booby trap, created by the loose lexicology of the word, *system.* Resort to Webster confuses rather than clarifies, with a generous offering of some seventeen definitions from which to choose. There are many more. Every writer on the subject seeks to achieve refinement of conceptualization through verbalization, a not uncommon ploy in the realm of the written word. Generally accepted ingredients of a working definition may be found in the following [1][2]: "a set of interrelated components working together to accomplish a certain objective." Not a bad statement, this. But not very useful, either, for it can be applied with equal appropriateness to my toaster and to TWA. Herein lies the making of another booby trap, that of fallacious analogy, for to impute verisimilitude where only superficial similarity exists leads to nonsensical conclusions about everything, including skill transfer. In this instance, to the fixit man who tinkers with my toaster will be attributed the capability of coping with TWA, its fleet of aircraft, its worldwide ticketing procedures, its luggage transfer and tracing mechanisms, its pilot and personnel training practices, and whatever of these singularly or collectively may come under the heading of *system.*

Not intended as mere pedantry, this example typifies the slippery semantic basis on which the "systems approach" has been rationalized as the panacea for society's ailments, past and future. Here at long last, we are led to believe, is the end of the by-guess-and-by-gorry era; horse-and-buggy methods are passé. We have the systems approach to win wars, design strategies for peace, curb crime, devise transportation, improve managerial efficiency in public and private sector alike, and guarantee a braver new world for the future. In our era of advanced technology, we have "Space Age techniques," as those associated with and derived from the systems approach are often called. Much like *jinn* (defined by Webster as "rational, but intangible beings"), these are the tools presumably appropriate to our every task and ready to serve our purposes.

## Development of Engineering as a Profession

That the engineering community should condone, let alone encourage, so amorphous a state of conceptualization as that which exists with respect both to the techniques and to their application is an anomaly, particularly when one reviews the historical development of engineering. Ever since the Middle Ages, when engineering was regarded as a kind of "craft mystery,"—a combination of artistry, intuition, and trial-and-error—there has been a discernible trend toward differentiation and specialization. During the Renaissance, the making of engines for battle (hence the word *engineer*), the draining of mines, and building of roads and bridges called for specific skills. Associations and societies reflected and strengthened the push toward greater differentiation. In recent years, professionalization and specialization have advanced in tandem relationship, with proliferation of fields within fields occurring; thus, the advent of nuclear engineering and computer technology since World War II. Each classification has generated its own body of theory, its own analytical descriptions and idealized models. To judge from the meticulous articulation of requirements for academic degrees in each, "crossover" or "passing" has been virtually impossible. In practice, the electrical engineer does not presume to design bridges; the nuclear engineer does not concern himself with sewage treatment; the computer specialist constructs no dams. But, in the present popular proclivity for *social* systems work, the demarcations again become blurred. Engineers *en masse* are exhorted by the deans of their colleges to turn their talents to solving the amorphous mess known as social problems—to become "public technologists," "civil science experts" and to tackle health, education and welfare as their social duty.

With heritage couched in the military so cozily that a special term, *civil engineering,* had to be coined to describe activities in the civilian sphere, perhaps engineering may be suffering from a case of what psychiatrists call "collective unconscious guilt." Psychologists have identified conscious guilt among workers in munitions factories, and engineers employed in the California experiment, to be discussed later, expressed the desire to be able to tell the wife and kids what they did at work. Perhaps, on the other hand, the quest for social service is a manifestation of the current striving for relevance.

With commencement exercises exhorting greater social commitment and professors urging a deeper dedication to social values, even to the extent of their taking a kind of Hippocratic Oath [2], engineers are being pushed willy-nilly into new roles.

They are caught between the Scylla of "doing their own thing well," which, in its pathological extreme has earned the label of "Eichmannism," and the Charybdis of abandoning their specialized capabilities in favor of a generalized doing of the public good. If the proposed code of ethics is to serve a useful purpose, its details must be spelled out. Otherwise, the implication would be allowed to stand that engineers have within their power the capacity, that the state-of-the-art has advanced to the point, and that all that is needed is a moral commitment, for them to solve social problems. Perhaps ethical dedication could more properly take the form of more mundane and less esoteric matters. Here is a random list of activities for engineers seeking "social relevance": Medical electronics not so faulty that they deliver death while monitoring [3], overpasses that do not cave

[1] Research Sociologist, Space Sciences Laboratory, University of California, Berkeley, Calif. Author of *Systems Analysis in Public Policy,* University of California Press, 1972 and 1973.

[2] Numbers in brackets designate References at end of Forum.

in, television sets not hazardous for electric shock potential, cars that perform safely even without the vigilance of Nader's crusaders.

Redundantly but inexorably, the breaking down of time-honored boundaries has come about through abdication of the very specificity and precision that formerly made for professionality. Through the conduit of "systems," which, variously defined, are the stock-in-trade of all branches of engineering, a kind of amalgamation has occurred. They all deal with systems; hence, they produce "systems experts." And with systems loosely defined to fit everything from the aforementioned toasters to TWA, the circular semantic syllogism is complete.

### Systems Analysis as Panacea

It behooves us to ascertain how this came about and where it is likely to lead. As to the first question, the phenomenon of receptivity toward systems techniques as a nostrum for all of society's ailments is not pure happenstance. Since as historiographers inform us, ours is the Computer Age, the Scientific Civilization, or the Space Age, then consistent with our *Zeitgeist* [4] is the notion that technological solutions are the ones best suited to solving our problems, the more so, it is argued, because a preponderance of these problems can be attributed to or are associated with technological causes. And since enormous and growing complexity is an outstanding characteristic of our web and tangle of social problems, only systems analysis, with its capacity to embrace the whole, order the parts, and achieve "rational" goals with mathematical precision is adequate to the task at hand.

But systems analysis has more than a temporal relationship with the Space Age; systems analysis epitomizes the Space Age. When the overworked cliché, "A nation that can put a man on the moon can ...," is evoked to prove technical prowess, reference is being made, wittingly or unwittingly, directly or indirectly, to the systems methods used with singular success in space exploration and, previously, under the name of *operations research*, in the Battle of Britain. Adding to the aura of invincibility was Secretary of Defense McNamara's astute use of the techniques in convincing Congress that war was too big business for the generals. With the Whiz Kids and the Rand Corporation to bedazzle and bemuse an unsophisticated Congress, the "scientific management" of procurement and strategy was added as a triumphant chapter to the saga of systems analysis. That the wizardry was short-lived, that subsequent Congressional investigations revealed fatal weaknesses in the techniques as management tools for the business of war are facts that have been cavalierly overlooked. A legend once born, a mythology was in the making. And those who could have penetrated the miasma swathing the high priests chose instead to take personal advantage of it. McNamara's band spun out to all branches of government and, in their endeavors, were joined by a heterogeny of professionals, paraprofessionals, and pseudoprofessionals—all bent on solving society's problems "scientifically and rationally." No one elected to keep the emperor honestly naked. And so systems analysis and its counterparts and components, management information systems, cost/benefit measures, program budgeting, performance and production evaluations, and the like have pursued a chameleonlike course, appearing in one guise or another in every branch and at every level of government. And, at every stage of this numbers game, there are "expert technicians," eager to desert their wonted fields for greener pastures. Engineers may not always be in the vanguard but their tools are!

In its journey from strictly engineering tasks and texts to the realm of the military and outer space, systems analysis has undergone considerable transformation. Lost is the methodological rigor sought in its original state; what remains are the empty forms, devoid of content. Jingo and jargon substitute for precision, the more so as systems analysis becomes the stock-in-trade of the motley assemblage of practitioners now claiming expertness. Besides the many kinds of engineers, there are economists, urban planners, policy analysts, operations researchers, management scientists, econometricians, and even sociologists. Engineering firms, computer manufacturers, and assorted research institutes, and "think tanks" are all busy peddling what has come to be known as "knowledge technology."

Perpetration and perpetuation of a hoax need not, however, be inferred as the reasons for the continuing bastardization of what were respectable and legitimate concepts in their own discipline. Systems analysis has come to represent the "dominant paradigm," to use Kuhn's term [5], of our time. (Kuhn describes the concept as a fundamental way of perceiving, thinking, and doing, consistent with a particular vision of reality.) Acceptance of the paradigm, which, Kuhn says, gains status if it has been demonstrated to solve *some* kinds of urgent problems, affords it a kind of legitimacy. It becomes a kind of language. Systems analysis can therefore be considered the dominant paradigm of our era. The prevailing mental set is such that approaches with technical trappings, that feature quantitative measures, and which sound "scientific" enjoy a hospitable environment. So welcome was the transplant of the tools from technical to social matters that the California experiment, conspicuous for its pioneering in this respect, became an historic landmark, criticism of which smacked of unpatriotic, perhaps even subversive, opposition to "scientific progress."

### The California Experiment

The original four contracts won plaudits without proving anything more decisive than the power of politics and salesmanship. Systems analysis was the vehicle designated by then Governor Brown to convey Space Age techniques to the public arena, where he wanted them utilized to design systems of criminal justice, information-handling, waste management, and mass transportation. Brown stressed growing magnitude and complexity of such public concerns as the factors rendering them amenable to systems analytic treatment. Improved governance was not his only objective. There was also the hope that the public sector would provide employment opportunity and avenues for diversification for skilled manpower threatened with job loss due to retrenchment in California's aerospace and defense industries. Although committees reviewing the proposals, submitted by an array of accounting firms, research institutes, computer manufacturers, and aerospace companies insisted that the sole criterion was merit, selection of the latter for the four original contracts may not have been due to pure chance, nor to undisputed superiority over all other contenders. The fact that Lockheed, North American Aviation, Aerojet General, and Space General were chosen lent initial credence and impetus to their subsequent campaign to dominate a market that was not yet displaying its full potential for bullishness.

Inquiry as to the qualifications of the engineers assigned to carry out the systems analysis of crime, transportation, etc., discloses a paradox. Many of them were chosen, not because of their demonstrated engineering capability but because of a college minor in business administration or a course in sociology. With only a few exceptions, they were probably not the companies' most valued engineers but, perhaps, the most expendable. The tendency to use "public technology" as a dumping ground for surplus engineers has become even more apparent as traditional jobs in industry have dried up. Public concern has been aroused over this new kind of unemployment; in recent years, increasing joblessness among midcareer engineers in their middle years and in middle income brackets has begun to cause uneasiness not only over the obvious economic consequences but about the more subtle possibility that long-cherished American myths would be undermined. Confidence in higher education as a guarantor of careers is shaking. Threatened are the neat correlations between unemployment and poor education, lack

of skill, and paucity of experience. To counteract a situation not statistically significant in total labor force numbers but unsettling as a sociological phenomenon, the League of Cities, the U.S. Department of Labor, and other agencies jointly sponsored programs to facilitate the "reorientation" of aerospace engineers. These were built on the assumptions that *here* was a reservoir of "systems capability" and that *there* was an accumulation of ailing "social systems" and that all that was needed were the will and way to make a match.

All three were in error. To begin with, many of the displaced engineers had not been in systems work. In fact, national statistics show that only a minute proportion have. In the second place, although public management problems abound, labelling them "civil systems" or "public service systems" does not render them readily amenable to the engineer's treatment. In the third place, despite the will, which sometimes displayed the force of a juggernaut, and the way, which manifested itself in federal funds under the Public Employment Program (PEP), the marriage did not last. The honeymoon was of short duration.

The engineers soon discovered the seriousness of the first booby trap mentioned earlier, i.e., the vagueness of the word, *system*. For them a system was a thing, its working was governed by certain laws; its actions were controllable and predictable. In the realm of social affairs, the system is an *idea;* it is a construct. In the words of a political scientist, "all systems are human artifacts designed to facilitate choices . . . They are overwhelmingly dependent on the hierarchy of values in the mind of the system's architect . . . To adopt such systems as 'real' implies a belief in the 'naturalness' of an imperfect heuristic device [6]." Only the least important of a system's workings are subject to immutable laws; if its actions are controllable and predictable in the short run, long run behavior is all the more likely to defy precise calculation. Face-to-face in a real-life situation, the engineers did not know what to do and the public agencies did not know what to do with them. For the engineer accustomed to explicit tasks performed in a highly structured situation, the amorphous situation was unsettling. For the public official, the specification of managerial objectives was tantamount to the demand for a quantified *apologia pro sua vita*, and in many cases he was not able to provide it. Many of the positions created under PEP petered out when the funds were exhausted.

### The Case of Public Welfare

Let us consider as a case in point public welfare, so often chosen as a target for systems treatment. How does one define a "public welfare system"? Welfare is mainly a philosophy, a social commitment of concern for the less fortunate. It is a set of activities. It is a complex of history and laws that vary as to geographical location and bureaucratic level. It is a tangle of relationships—between people, between conflicting and overlapping agencies, between intangible forces. It is a political whipping-boy, the brunt of blatant manipulation by officials who use its pliable statistics to suit their purposes. What are its objectives? The requests for proposals for systems studies stress "reduction of dependency." This would imply that welfare systems should be designed to get people off them. But one wonders whether that objective should not more properly be the responsibility of *other* systems in the society. If the education system supplied sufficient skills for success in jobs; if the labor market system behaved so as to provide adequate employment opportunities and pay; if there were a system of child care that would enable women to work; if public health systems operated in such fashion that illness would not be financially catastrophic; if public housing systems functioned so that people could live safely and decently even on low incomes—would not these be more socially desirable and more efficacious than a public welfare system "to reduce dependency"?

The engineers, faced with the frustration of reducing to a simplistic model the complexities of social welfare, resorted to the busy work of information-gathering. Observation of innumerable systems studies indicates that this step, indicated as part of the larger analysis, often engulfs the whole effort. As a result, a systems analysis of any given system becomes preoccupied with its information as though that were an end in itself. For the engineers working on public welfare, development of an information system provided an activity which they found *gemütlich*. And, with their *Weltanschauung*, the objective of such a system was to save money by screening out all the bad guys, bums, freeloaders, and welfare cheats that they assumed must be the cause of the rising costs. What they designed was a system so expensive that its operations would have constituted a severe and permanent drain on the State, probably far exceeding the cost of maintaining the fraudulent cases that exist and never coming close to the big abuses countenanced by the political aspects of public welfare, nor to the realities involved in welfare costs.

The engineer soon discovers that crime, welfare, and the like are not one system, but many systems; not one problem, but many problems, only some of which fall within any disciplinary or "scientific" scope. Indeed, for most social problems, there is no technical or methodical "solution." Their very definition depends on the point of view of the analyst; his value system is imbedded in the way he designates the objectives, identifies pertinent variables, and assigns costs and benefits. And in all these crucial matters, his schooling and experience as an engineer may hinder rather than help him. Thanks to the process of "trained incapacity," he, probably more than any other professional, purposefully rivets attention on certain types of problems or aspects of them and does well what he does, with careful disregard for those matters which he considers outside his purview.

### Socialization of the Engineer

With no pejorative intent, the point must be made that engineers, probably to a greater degree than any other group, are caught in the dilemma of social responsibility. Ethical issues have been considered by their professional associations only insofar as they have affected the relations of individual engineers with their colleagues and employers but rarely with the complex and diffuse questions relating to the society at large [7]. Professor William M. Evans, who has conducted authoritative research into this matter, offers the following theory:

> One reason for the widespread neglect on the part of engineers of the problem of social responsibility for technological change is the difficulty of accepting responsibility for events *over which they exercise no control*. As salaried employees performing in the main a staff function, engineers are rarely in a position of making policy decisions concerning the wisdom of developing or not developing a new engineering *product, or concerning* what, if any, action might be taken to counteract its potential or actual negative social effects. This is particularly true for overwhelming proportion of engineers engaged in production or in development research, where the norms governing their conduct are neither *service*—typical of the professional-client relationship—nor *autonomy*, typical of the scientist-employer or in the science-patron relationship, but *obedience*, typical of the employee in the employer-employee relationship [8].

From C. P. Snow [9] to Seymour Lipset [10] with Smithers [11] in between, we can, in addition, trace a pattern of career preferences, political perspectives, and private aspirations that characterize engineers as possessed with "a basic lack of sociability" and a strong incidence of "low people-orientation."

Engineers, like all other forayers into unwonted places, demonstrate the validity of Abraham Maslow's sage maxim: "If the only tool you have is a hammer, you tend to treat everything as if it were a nail [12]." With the systems approach the prevailing paradigm of our time, the tendency is to construe all of society's ailments as though they were malfunctioning "systems," and in applying the hammer treatment, engineers, for reasons historical, political, and economic, constitute the core of latter-day social saviors. They have been designated the standard-bearers to "bring science and engineering, through technology, to bear more effectively on societal problems." In all fairness, therefore, we must take care lest we kill the messenger who brings bad tidings. Perhaps it is not the engineers who should be criticized for behaving true to form, in only doing what is asked of them. Perhaps we should look farther and direct our criticism to government agencies which should know better but which still perform the rites of the Cargo Cult [13]. This is the current activity of the prestigious and respected National Science Foundation and the General Accounting Office. The former, in a remarkable report, noteworthy for its ignorance of the performance of "the highly promising, sophisticated new methods . . . particularly systems analysis, computerization, and the mathematical simulation in 'models' of the dynamics of complex problems [14]," is called upon to "formalize the participation of scientists or engineers" in "seeking solutions to the major problems confronting our people." Specifically, under terms of H. R. 15789 (July 1972) introduced by Rep. John Davis, a member of the House Committee on Science and Astronautics, there would be established a "Civil Science Systems Administration," with $25 million for 1973, $20 million for 1974 and $10 million for 1975. The descending scale presumably indicates that the national problems pinpointed as "health care, poverty, public safety, pollution, unemployment, housing, education, transportation, nutrition, communications, and energy resources," responded overtime to the treatment. Out of this legislation there emerged "Public Technology, Inc.," developer of "Urban Technology System" and dedicated presumably to carrying out the Presidential mandate, high-minded and totally unworkable, used as a preface for the NSF report [15]:

> My colleagues in the Congress, America can be her true self only when she is engaged in a great enterprise.
>
> To build a full generation of peace is a great enterprise.
>
> To help the poor and feed the hungry, to provide better health and housing and education, to clean up the environment, to bring new dignity and security to the aging, to guarantee equal opportunity for every American—all these are great enterprises.
>
> To build the strong economy that makes all these possible—to meet the new challenges of peace, to move to a new prosperity without war and without inflation—this truly is a great enterprise, worthy of our sacrifice, worthy of our cooperation and worthy of the greatness of a great people.

With disregard for the flaws of the systems approach in the management of the military establishment, its own home territory, the General Accounting Office furthered the cause of "great enterprises" by generating a report to Congress on "Means for Increasing the Use for Defense Technology for Urgent Public Problems" [16]. Parroting the popular verbiage, instead of providing the critical inquiry more typical of its public service, the G.A.O. explains why the inquiry has been undertaken:

> The Congress and the President have expressed concern that greater application be made of the vast scientific and technological resources of the Federal Government—until now directed in large part to defense and space problems—to help solve massive sociological problems facing the United States.

Designated as target areas in *this* context are transportation, law enforcement, environmental protection, education, housing and urban development. The wishful thought embodied here is that this effort will yield the following benefits: "nondefense markets should evolve and the conversion could have a favorable impact on employment, the domestic economy, and the international trade balance" [17]!

The roster of problems grows in length and in complexity. Does the state-of-the-art advance commensurately? I think not. Even the cliché that caused so much of the cult has now been called into question. The nation that could put a man on the moon could insure the technical performance of the hardware but created havoc in personal lives and emotional upheaval beyond repair. Life stories of astronauts subsequent to their return to terrestrial pursuits attest to the damage, recondite but real, incalculable but important, caused by these exploits [18]. Mathematical precocity has saddled us with outrageously costly programs that assure neither peace nor prosperity; technical virtuosity continues to bedazzle the uninitiated and to subvert social planning. We have been hoisted on our own petard. We welcomed technical solutions because ours is a technological era, because we interpreted our problems as being technological, and because we imputed to technology the capability of solving them.

### Systems Analysis as Technique for Solving Social Problems

If we had examined the techniques divested of all the rationalizations, we might have recognized earlier that they simply were not appropriate to the myriad of social tasks to which they were being assigned. In order for the desired technological transfer to take place, certain conditions had to be met and certain assumptions about systems made explicit. In engineering, the application of systems analysis to a problem is predicated on knowledge of a family of parameters that describe and define that system in precise terms. Built in here is explicit recognition of a set of values of those parameters at which the system performs optimally [19]. In social planning, there exist neither undisputed parameters nor uncontrovertible criteria. Quite the contrary. These lie only in the eye of the beholder, and thus are subject to the most biased interpretations and manipulation. The evidence indicates that the "new utopia" [20] created by engineers may not be the best of all possible worlds.

Moreover, some of the piles on which the case for technology transfer has depended have been systematically demolished by careful and methodical analysis. Professor L. A. Zadeh of the Electronics Research Laboratory of the University of California's College of Engineering contends that the capacity for dealing with complexity, the main selling point for systems analysis as practiced in engineering, just does not exist. Instead, as the complexity of a system increases, "our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost *mutually exclusive characteristics* (emphasis added). It is in this sense that precise quantitative analyses of the behavior of humanistic systems are not likely to have much relevance to the real-world societal, political, economic and other types of problems which involve humans whether as individuals or in groups [21]." From his observations, Zadeh has developed what he calls the *principle of incompatibility*, to describe the intrinsic unsuitability of the conventional quantitative techniques of systems analysis for dealing with complex real-life social problems.

The incompatibilities are underscored by Robert A. Frosch

(then Assistant Secretary of the Navy) in a series of antitheses that "pit the systems world against some aspects of the real world [22]." In so doing, Dr. Frosch disencumbers the loaded argument that the *approach* is good and that only the uses are faulty.

> To the charge that I am writing about bad systems engineering, I can only say that I am taking a pragmatic view. The thing is defined by what is done, not what is said; and if what I am describing is bad systems engineering, I can only say that I seldom see any other kind.

He demonstrates that these techniques are proving to be—

> a poor substitute for real science and engineering. They are at the least, not doing what they are advertised as doing, if they are not *making things worse*. It could be that things would be even worse without these new techniques, but I would like to ask some questions and suggest some reasons for believing that systems engineering, systems analysis, and management, as practiced, are likely to be part of the problem, and indeed causative agents.

We consider ourselves to be a rational people. We are comfortable with the simplistic simulations of the first step in systems analysis and choose to ignore the implications to Zadeh's and Frosch's contributions. Our technologically bemused society might do well, in this regard, to heed Banquo's observations on the witches' performance:

> And often times, to win us to our harm
> The instruments of darkness tell us truths,
> Win us with honest trifles, to betray us in deepest consequence [23].

### References


1 Borrowed from C. West Churchman.
2 *New York Times*, 10-9-73, "Professor Urges Oath for Engineers."
3 London, Michael P., "Defects in Medical Electronics Draw Heavy Fire from Hospitals," *Electronic Design*, Oct. 28, 1971, pp. 22–26.
4 Spirit of the times.
5 Kuhn, T. S., *The Structure of Scientific Revolution*, University of Chicago Press, Third Impression, 1971.
6 Haas, Ernest B., and Ruggie, John G., "International Technology and International Action: What, Who, How and Why?" Paper issued by the Institute for International Studies, Technology Utilization and International Systems Project.
7 National Society of Professional Engineers, *Ethics for Engineers; Canons of Ethics, Creed & Rules of Professional Conduct*, Washington, D.C., National Society of Professional Engineers, 1962.
8 Evans, William M., "The Engineering Profession: A Cross-Cultural Analysis," from a chapter in Robert Perrucci and Joel Gerstl, eds., *The Engineers and the Social Systems*, New York, John Wiley (in press).
9 Snow, C. P., *The New Men*, London, Macmillan, 1954.
10 Lipset, Seymour M., and Ladd, Everett C., "Politics of Academic Natural Scientists and Engineers," *Science*, Vol. 176, June 9, 1972.
11 Smithers, Alan, "Occupational Values of Students," *Nature*, Vol. 222, May 24, 1969.
12 Maslow, Abraham, *The Psychology of Science: A Reconnaissance*, New York, Harper & Row, 1966.
13 The Cargo Cult is a tribal custom practiced in the Melanesian Islands by natives who, having seen "silver birds" laden with treasure (cargo) materialize out of the sky after U.S. Navy "Sea Bees" cleared the jungle for runways during WW II, still hack out the jungle and expect the "silver birds" to return.
14 Report of the National Science Board, *The Role of Engineers and Scientists in a National Policy for Technology*, National Science Foundation, 1972, p. 5.
15 President Richard M. Nixon, before a Joint Session of the Congress, September 9, 1971.
16 Comptroller General of the United States, *Means for Increasing the Use of Defense Technology for Urgent Public Problems*, General Accounting Office, B-175132, Dec. 29, 1972.
17 Ibid., p. 12
18 Aldrin, Edwin E., *Return to Earth*, 1973.
19 Silver, Samuel, "Science and Technology," *Science Journal*, *5A*, No. 4, Oct. 1969, pp. 39–44.
20 Boguslaw, Robert, *The New Utopians*, 1965.
21 Zadeh, L. A., "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," Memorandum No. ERL-M342, University of California, College of Engineering, Electronics Research Laboratory, July 24, 1972, pp. 2–3.
22 Frosch, Robert A., "A New Look at Systems Engineering," IEEE *Spectrum*, Sept. 1969, pp. 24–28.
23 Shakespeare, Wm., *Macbeth*, Act I, Scene IV, 1, pp. 123–127.


## Numerical Integration of Stiff Dynamic Systems


**L. LAPIDUS[1] AND R. C. AIKEN.[1]**


Many commonly occurring dynamic systems have widely separated time constants. These systems are often represented by sets of initial value ordinary differential equations which possess some variables that rapidly change during time intervals much smaller than the transient phenomenon of interest. Such systems have "stiff" stability requirements for numerical integration. It is of major importance to communicate to the uninitiated how to recognize the stiff system and effectively proceed to calculate the numerical solution. Otherwise, confusion can arise as to why the commonly available integration routines require inordinately small step sizes to obtain only an insignificant fraction of the total transient, at great expense in computer usage.

Examples of such systems abound in engineering science. Control processes may be stiff as a result of fast electrical control circuits coupled to slow mechanical movement [1–3].[2] Electrical circuits may contain within the network semiconductors with very small time constants compared with slower high resistance elements [4–6]. Energy component contributions may be quite different in aircraft and rocket flight [7–9], exhibiting stiff characteristics. In chemical kinetics the stiff variables correspond to intermediate species that quickly react or are present in low concentrations [10]. Enyzme kinetics [11], polymerization [12], and chain reactions [13] illustrate this feature. The stiffness of these reactions remains when coupled with physical transport processes in an industrial reactor [14].

The step size ($h$) limitations for most standard algorithms on an $n$-dimensional system require $\max|h\lambda_i|$, $i = 1, 2, \ldots, n$, where $\lambda_i$'s are local eigenvalues, bounded by a single small number, typically in the 1 to 10 range. If any eigenvalue is large in absolute value, corresponding to a fast mode, severe restrictions are placed on the step size. Depending on the length of the solution interval of interest, this may demand a great deal of computation time. Further, there are limits on how small $h$ can be before roundoff errors accumulate and render this calculation worthless.

Although any contribution to the solution by the stiff variable occurs only during a small portion of the transient (the "boundary layer"), determined by $\max|\lambda_i|$, this eigenvalue will continue to fix the manimum allowable stepsize throughout the domain of interest. It is particularly important to specify this domain in defining a stiff problem, and this is often overlooked. Thus for some applications the boundary layer may be of primary con-


[1]Department of Chemical Engineering, Princeton University, Princeton, N. J.


[2]Numbers in brackets refer to references at end of Forum.

D6

VIROLOGY 60, 485–490 (1974)

# Correlation Between Amino Acid Exchanges in Coat Protein of TMV Mutants and the Nature of the Mutagens

B. SINGER AND H. FRAENKEL-CONRAT

*Space Sciences Laboratory and Department of Molecular Biology and Virus Laboratory, University of California, Berkeley, California 94720*

*Accepted March 26, 1974*

This paper represents a comparative study of the action of various reagents of different mutagenicity on the messenger activity of TMV RNA, as monitored by amino acid analyses of the coat protein of 225 mutants. No such comparative data appear to be available from other laboratories.

The stable mutants with single amino acid exchanges obtained with nitrous acid, hydroxylamine, and methoxyamine showed the expected predominance of those changes which would be predicted on the basis of their reaction specificity (C → U and A → G; C → U).

The amino acid exchanges caused by bromination and alkylation indicated a greater variety of mutagenic events. Recent studies of the many reactions caused by alkylating agents, indicating that all bases may be modified at more than one site, supply an explanation for this finding.

The results of photochemical reactions which destroy bases resemble those of spontaneous mutations in giving a completely random exchange pattern.

## INTRODUCTION

The studies of the chemical and the mutagenic effects of nitrous acid acting on TMV-RNA represented the first proof that point mutations are due to nucleic acid base modifications (Mundry and Gierer, 1958; Schuster and Schramm, 1958; Schuster and Vielmetter, 1961). The specific effects of many different chemicals on polynucleotides have since been intensely studied. It has become evident that the predominant reaction need not represent the biologically effective one. Particularly with alkylating reagents minor reactions have recently been discovered that may actually be responsible for the mutagenic and/or carcinogenic efficiency of the respective reagent.

In our laboratory we have been conducting parallel studies on the chemistry of many mutagens with particular reference to the formation of minor derivatives, and also mutating TMV or TMV-RNA with these same reagents and measuring the mutation frequency, and analyzing the mutant coat proteins. We now present the results of amino acid analysis of 225 mutants of TMV obtained with a variety of modifying reactions, and the attempts to correlate the observed amino acid exchanges with what is presently known of the chemistry of these reactions.

## MATERIALS AND METHODS

TMV or its RNA, depending on the reagent, was reacted with $HNO_2$ according to Mundry and Gierer (1958); hydroxylamine (HA) and methoxyamine (MA) according to Fraenkel-Conrat and Singer (1972); nitrosoguanidine, dimethyl sulfate, and methyl methanesulfonate according to Singer and Fraenkel-Conrat (1969a,b); and aqueous bromine or *N*-bromosuccinimide according to Brammer (1963) and Means and Fraenkel-Conrat (1971).

The photocatalyzed reactions of TMV-RNA with iron salts or dyes were according to Singer and Fraenkel-Conrat (1965,



Copyright © 1974 by Academic Press, Inc.
All rights of reproduction in any form reserved.

1966). Very brief treatment of TMV-RNA with snake venom phosphodiesterase was according to Singer and Fraenkel-Conrat (1963). The chemical elimination of the terminal nucleoside of TMV-RNA was according to Steinschneider and Fraenkel-Conrat (1966).

Mutational reactions were generally performed on TMV-RNA which was then reconstituted with TMV protein (Fraenkel-Conrat and Singer, 1959) before inoculation to plants. Nitrosoguanidine is highly mutagenic only when TMV-RNA is in the virus, and for this reason only intact TMV was used.

The method for the determination of infectivity and mutation frequency are given by Singer and Fraenkel-Conrat (1969c). Tables of relative mutagenicity have been published by Singer and Fraenkel-Conrat (1969c,d).

The usual methods were used for obtaining purified preparation of TMV mutants by differential centrifugation and of the protein by the acetic acid method (Tsugita and Fraenkel-Conrat, 1962). The techniques of hydrolysis and amino acid analysis were also the same as in earlier studies. Analytical data on isoleucine, leucine, and valine were based on 72-hr hydrolysis, and those for serine and threonine on extrapolation to 0 time from the 24 and 72 hr analytical data.

## RESULTS

The method of selection of mutants for study was, for practical reasons, the isolation of those giving local lesions on *N. sylvestris*.[1] The process of freeing such mutants from wild-type virions was by repeated transfer to the same host and/or to *Nicotiana tabacum* var. Xanthi nc. Only a small percentage of the mutants obtained by chemical treatment of TMV or its RNA survived such repeated transfers, which is in distinct contrast to the spontaneous mutants or contaminants present in untreated TMV, which were completely stable during such repeated transfers. About 10% of the viable mutants derived from chemical modification were unstable in the sense that they changed in symptomology and/or amino acid composition upon repeated transfers. These are believed to represent products of secondary spontaneous mutations of low viability mutants, and they are grouped together with previously discussed mutants differing grossly from wild-type in coat protein structure and resembling other TMV-groups (e.g., group B = Dahlemense, etc.) (Tsugita and Fraenkel-Conrat, 1962; Tsugita, 1962).

Of the 197 stable mutants here reported on 12 showed exchanges of 2 or 3 different amino acids and these are not included in the analysis of the data, because their codon assignment is not generally unambiguous. One third of the remaining mutants showed no amino acid exchanges in the coat protein; these presumably carry exchanges in other, yet undetermined, virus-coded proteins.

The observed amino acid exchanges in the remaining 117 mutants, and the frequencies of these exchanges as obtained after various mutagen treatments, are listed on Table 1. Most of these, but particularly the unexpected ones (other than C → U or A → G) and those giving ambiguous or unstoichiometric analyses (reproducibly $\gtrsim 0.25$ above or below an integer number of amino acid residues per mole) were regrown and the protein of the progeny virus was reanalyzed 2–4 times, a procedure which led occasionally to the detection of unstable mutants, as discussed above and not included in the table. Table 2 summarizes all the definitive data. The frequencies of obtaining exchanges other than C → U or A → G are listed on Table 3.

Amino acid exchanges resulting from nitrous acid mutagenesis can be accounted for to 89% by C → U and A → G modifications, in good agreement with earlier data from the laboratories of Fraenkel-Conrat and Wittman (summarized by Funatsu and Fraenkel-Conrat, 1964), and in support of previous conclusions that the deamination of cytosine and adenine leads to miscoding, but not that of guanine (Schuster and Vielmetter, 1961). Hydrox-

[1] Earlier studies from this laboratory included mutants which differed only in systemic symptomology from wild type TMV. These mutants are not included in the present summary.

TABLE 1

FREQUENCY OF AMINO ACID EXCHANGES IN COAT PROTEIN OF PROGENY VIRUS OF TMV TREATED WITH VARIOUS MUTAGENS

| Amino acid exchange[b] | Possible codon exchange | Modification reaction[a] | | | | | |
|---|---|---|---|---|---|---|---|
| | | $HNO_2$ | HA, MA | Alkyl | $Br_2$ | Photo chem. | Term. elim. |
| Ala → Ser | G → U | | | 1 | | | |
| Ala → Val | C → U | | | 2 | | | |
| Arg → Gly | A → G | 5 | | 1 | 1 | | |
| Arg → Leu | G → U | | | | 2 | | |
| Arg → Lys | G → A | 1 | | | | | |
| Asn → Ser | A → G | 5 | | | 1 | | |
| Asp → Gly | A → G | | | 1 | | | |
| Gln → Arg | A → G | 1 | | | | | |
| Gln → His | Pu → Py | | | | | 4 | 3 |
| Glu → Gly | A → G | 1 | | | | | |
| Ile → Asn | U → A | | | | | 1 | 1 |
| Ile → Thr | U → C | | 1 | 4 | 2 | 1 | 2 |
| Ile → Val | A → G | 4 | 1 | 4 | 2 | 1 | 2 |
| Leu → Thr | 2 exch. | | | 1 | | | |
| Pro → Leu | C → U | 4 | 1 | 3 | 5 | | |
| Pro → Ser | C → U | | | 1 | | 2 | |
| Ser → Gly | A → G | | | | 1 | | |
| Ser → Leu | C → U | | | 1 | | 1 | |
| Ser → Phe | C → U | 9 | 7 | 7 | 3 | | 1 |
| Ser → Pro | U → C | | | | | | 1 |
| Thr → Ala | A → G | 1 | | | | | |
| Thr → Ile | C → U | 2 | | | | | |
| Tyr → Phe | A → U | | | 1 | | | |
| Val → Ala | U → C | | | 1 | | | |
| Val → Ile | G → A | 2 | | | | | |
| Val → Met | G → A | 1 | | 1 | | 1 | 2 |
| Val → Phe | G → U | | | 1 | | | |
| Val → Pro | 2 exch. | | | | | 1 | |

[a] The modification reactions are abbreviated as follows: nitrous acid, $HNO_2$; hydroxylamine and methoxyamine, HA and MA; alkylation including nitrosoguanidine, dimethyl sulfate, and methyl methane sulfonate, Alkyl; *N*-bromosuccinimide and bromine water, $Br_2$; dye and metal catalyzed photochemical reactions, Photo-chem; brief snake venom phosphodiesterase treatment or stepwise chemical elimination, Term-elim.

[b] In the cases of exchanges involving aspartic or glutamic acids or their amides, the one accounting for the observed exchange on the basis of a single nucleotide replacement is arbitrarily listed.

ylamine and methoxyamine primarily substitute the exocyclic amino group of cytosine in ribopolynucleotides (Fraenkel-Conrat and Singer, 1972), enabling it to pair like uracil (Brown *et al.*, 1968; Singer and Fraenkel-Conrat, 1970), and the observed amino acid exchanges are in good accord with this mechanism (Table 1).

The effects of alkylating agents on RNA and DNA is actively being studied in several laboratories, and is reviewed by Singer (1974). There is substantial agreement that the predominant reaction, the 7-alkylation of guanine is not responsible for the carcinogenic effects of many alkylating agents (Swann and Magee, 1968, 1971; Schoental, 1969; Ross *et al.*, 1971; Lijinsky *et al.*, 1972; O'Connor *et al.*, 1972; Kleihues and Magee, 1973; Craddock, 1973) and is not per se mutagenic (Ludlum, 1970). In the specific case of nitrosoguanidine acting on intact tobacco mosaic virions, methylation of cytosine could be correlated with mutagenicity (Singer and Fraenkel-Conrat, 1969b,c, 1970). On the other hand, it is suggested by recent data

TABLE 2

SUMMARY OF MUTANTS OF TMV PRODUCED WITH VARIOUS MUTAGENS AND CODON EXCHANGES DERIVED FROM AMINO ACID ANALYSIS OF THEIR COAT PROTEIN

| Modification reaction[a] | Number of mutants[b] | | | | | Codon exchanges[c] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Unstable | Multiple exchanges | No exchange | Single coat protein exchange | Number | | | % | | |
| | | | | | | C → U | A → G | Other | C → U | A → G | Other |
| $HNO_2$ | 65 | 11 | 7 | 11 | 36 | 15 | 17 | 4 | 42 | 47 | 11 |
| HA, MA | 23 | 2 | 1 | 10 | 10 | 8 | 1 | 1 | 80 | 10 | 10 |
| Alkylating agents | 73 | 9 | 2 | 25 | 30 | 14 | 6 | 10 | 46 | 20 | 33 |
| $Br_2$ | 28 | 4 | 1 | 6 | 17 | 8 | 5 | 4 | 47 | 29 | 24 |
| Photochemical reactions | 20 | — | — | 8 | 12 | 3 | 1 | 8 | 25 | 8 | 67 |
| SV-DE or term. chem./elim. | 16 | 2 | 1 | 1 | 12 | 1 | 2 | 9 | 8 | 17 | 75 |
| Total | 225 | 28 | 12 | 61 | 117 | 49 | 32 | 36 | | | |

[a] References for the reaction conditions are given in Methods and Materials section. Fifty-eight of the alkylated mutants were obtained with nitrosoguanidine treatment of TMV. The rest were obtained with dimethyl sulfate or methyl methane sulfonate treatment of TMV-RNA.

[b] The total figure refers to the number of separate mutants and does not include reisolated mutants. Unstable mutants are those which upon reisolation exhibit changed symptomology and/or changed amino acid composition. Multiple exchanges refers to more than 1 exchange, except when 2 identical amino acids are exchanged for two other identical amino acids (e.g., 2 Ser → 2 Phe).

[c] Codon exchanges are given only for those mutants listed in the category of "coat protein exchange," and do not include those with multiple exchanges (except as noted in footnote[b]) or unstable mutants. Codon exchanges listed as "Other" are given in detail in Table 3. All amino acid exchanges are given in detail in Table 1.

TABLE 3

FREQUENCY OF CODON EXCHANGES OTHER THAN C → U AND A → G[a]

| Codon exchange | Modification reaction[b] | | | | | | Total exchanges |
|---|---|---|---|---|---|---|---|
| | $HNO_2$ | HA, MA | Alkyl | $Br_2$ | Photochem. | Term. elim. | |
| Pu → Pu | | | | | | | 8 |
| G → A | 4 | | 1 | | 1 | 2 | |
| Pyr → Pyr | | | | | | | 12 |
| U → C | | 1 | 5 | 2 | 1 | 3 | |
| Pu → Pyr | | | | | 4 | 3 | 12 |
| G → U | | | 2 | 2 | | | |
| A → U | | | 1 | | | | |
| Pyr → Pu | | | | | | | 2 |
| U → A | | | | | 1 | 1 | |
| double exchange | | | 1 | | 1 | | 2 |

[a] Data from Table 1.

[b] Abbreviations are the same as in Table 1.

on alkylation of DNA and RNA with certain nitroso compounds that the substitution of the $O^6$ (O'Connor *et al.*, 1972; Lawley and Shah, 1972; Craddock, 1973) or possibly the 3 position of guanine (Lawley *et al.*, 1971) may be correlated with the high carcinogenicity of these compounds. The amino acid data obtained with alkylation mutants of TMV (Table 1), largely due to nitrosoguanidine treatment of the virus, since this is the only effective alkyl-mutagenizing method, support the predominant modification of cytosine. The relatively large number of C → U exchanges (Table 2), however, would not have been predicted, since none of the chemical effects of alkylating agents would be expected to reverse the pairing potential of the nucleotides, but rather render the base nonspecific, as demonstrated for 3 alkylcytosine (Singer and Fraenkel-Conrat, 1970; Ludlum, 1970, 1971; Ludlum and Magee, 1972), and $O^6$-alkylguanine (Gerchman and Ludlum, 1973).

Bromination of RNA causes primarily guanine and cytosine modification (Bram-

mer, 1963). However, only the reaction with cytosine has been studied in detail, and, when brominated poly(C) was used as template for RNA polymerase, it was found to pair with A as well as G, indicating that this reaction is mutagenic and that modified cytosine acts like uracil (Means and Fraenkel-Conrat, 1971). (Surprisingly, however, neither poly-5-bromocytidylic acid nor poly-5-hydroxycytidylic acid mispair in this system.) The data in Table 2 do show a preponderance of C → U exchanges in the coat protein of bromine-induced mutants. However, there are a number of other exchanges (G → U, U → C) (Table 3).

When RNA is illuminated with visible light in the presence of proflavin or thiopyronin, guanine is preferentially photodecomposed (Singer and Fraenkel-Conrat, 1966). Illumination of RNA in the presence of $Fe^{2+}$ or $Fe^{3+}$ leads to general destruction of bases (Singer and Fraenkel-Conrat, 1965). Neither the metal- nor dye-catalyzed reaction causes a high mutation frequency, but those mutants found would be expected to show a random pattern of amino acid exchanges and the data indicate that 6 of the 12 possible codon exchanges are found in the 12 mutants analyzed (Tables 2 and 3).

Occasional mutants with amino acid exchanges observed after treating RNA with the methods of stepwise degradation or by minimal snake venom diesterase treatment, both not expected to be mutagenic, showed a similarly random pattern, in support of the belief that these are spontaneous mutants.

## DISCUSSION

Nitrous acid is the only reagent that is able to effect a direct base change, C → U. The action of nitrous acid on adenine, and that of hydroxylamine or methoxyamine on cytosine lead to products that favor the alternate tautomeric state (hypoxanthine and $N^4$-hydroxy- or $N^4$-methoxycytosine) and thus are effective mutagens.

Alkylating agents are capable of a wide variety of base modifications besides the predominant 7-alkylation of guanine. None of these products of alkylation would be able preferentially to pair in opposite manner. The alkylation of the 7 position of guanine appears to be without biological consequence, except due to its tendency to become depurinated in DNA. The alkylation of the 3 position of cytosine and of the $O^6$ of guanine make these unable to form typical pairs of H-bonds. These two modifications have been shown to mispair in random fashion. The methylation of cytosine was shown to be mutagenic in TMV, and the methylation of the $O^6$ of guanine was suggested to account for the carcinogenic effects of certain nitrosoalkyl compounds. It has further been suggested that mispairing may be due to 3-methylguanine, 3-methylthymine, and $O^4$-methylthymine, all recently found to be minor products in nucleic acids treated with the carcinogen, nitrosomethylurea (Lawley *et al.*, 1971; Lawley and Shah, 1972; Lawley *et al.*, 1973). Certain ethylating agents, also found to be carcinogenic or mutagenic, have additional specificities, and substitute guanosine on the 1 position (Singer, 1972), and cytidine and adenosine on the exocyclic amino group (Sun and Singer, 1974; Singer *et al.*, 1974).

Alkylating agents, with their potential for modifying bases on many sites, none of which appear to be specifically mutagenic, might exert their powerful biological effects through multiple synergistic action.

## ACKNOWLEDGMENT

This investigation was supported by Grant No. CA 12316 from the National Cancer Institute and Grant No. GB 32367 from the National Science Foundation.

## REFERENCES

BRAMMER, R. W. (1963). *Biochim. Biophys. Acta* **72,** 217–229.

BROWN, D. M., HEWLINS, M. J. E., and SHELL, P. (1968). *J. Chem. Soc.* **C,** 1925.

CRADDOCK, V. M. (1973). *Biochim. Biophys. Acta* **312,** 202–210.

FRAENKEL-CONRAT, H. and SINGER, B. (1959). *Biochim. Biophys. Acta* **33,** 359–370.

FRAENKEL-CONRAT, H., and SINGER, B. (1972). *Biochim. Biophys. Acta* **262,** 264–268.

FUNATSU, G., and FRAENKEL-CONRAT, H. (1964). *Biochemistry* **8,** 1356–1362.

GERCHMAN, L. L., and LUDLUM, D. B. (1973). *Biochim. Biophys. Acta* **308,** 310–316.

KLEIHUES, P., and MAGEE, P. N. (1973). *J. Neurochemistry* **20,** 595–606.

LAWLEY, P. D., ORR, D. J., and SHAH, S. A. (1971). *Chem.-Biol. Interactions* **4**, 431-434.

LAWLEY, P. D., and SHAH, S. A. (1972). *Biochem. J.* **128**, 117-132.

LAWLEY, P. D., ORR, D. J., SHAH, S. A., FARMER, P. B., and JARMAN, M. (1973). *Biochem. J.* **135**, 193-201.

LIJINSKY, W., GARCIA, H., KEEFER, L., LOO, J., and ROSS, A. E. (1972). *Cancer Res.* **32**, 893-897.

LUDLUM, D. B. (1970). *J. Biol. Chem.*, **245**, 477-482.

LUDLUM, D. B. (1971). *Biochim. Biophys. Acta* **247**, 412-418.

LUDLUM, D B., and MAGEE, P. N. (1972). *Biochem. J.* **128**, 729-731.

MEANS, G. E., and FRANEKEL-CONRAT, H. (1971). *Biochim. Biophys. Acta* **247**, 441-448.

MUNDRY, K. W., and GIERER, A. (1958). *Z. Vererbungslehre* **89**, 614-630.

O'CONNOR, P. J., CAPPS, M. J., and CRAIG, A. W. (1972). *Biochem. J.* **129**, 519-528.

ROSS, A. E., KEEFER, L., and LIJINSKY, W. (1971). *J. Nat. Cancer Inst.* **47**, 789-795.

SCHOENTAL, R. (1969). *Proc. Biochem. Soc.* **114**, 55P-56P.

SCHUSTER, H., and SCHRAMM, G. (1958). *Z. Naturforsch.* **13b**, 697-704.

SCHUSTER, H., and VIELMETTER, W. (1961). *J. Chim. Phys.* **58**, 1005-1010.

SINGER, B.(1972). *Biochemistry* **11**, 3939-3947.

SINGER, B. (1974). *Progr. Nucleic Acid Res. Mol. Biol.*, in press.

SINGER, B., and FRAENKEL-CONRAT, H. (1963). *Biochim. Biophys. Acta* **72**, 534-543.

SINGER, B., and FRAENKEL-CONRAT, H. (1965). *Biochemistry* **4**, 226-233.

SINGER, B., and FRAENKEL-CONRAT, H. (1966). *Biochemistry* **5**, 2446-2450.

SINGER, B., and FRAENKEL-CONRAT, H. (1969a). *Biochemistry* **8**, 3260-3266.

SINGER, B., and FRAENKEL-CONRAT, H. (1969b). *Biochemistry* **8**, 3266-3269.

SINGER, B., and FRAENKEL-CONRAT, H. (1969c). *Virology* **39**, 395-399.

SINGER, B., and FRAENKEL-CONRAT, H. (1969d). *Progr. Nucleic Acid Res. Mol. Biol.* **9**, 1-30.

SINGER, B, and FRAENKEL-CONRAT, H. (1970). *Biochemistry* **9**, 3694-3701.

SINGER, B., SUN, L., and FRAENKEL-CONRAT, H. (1974). *Biochemistry* **13**, 1913-1920.

STEINSCHNEIDER, A., and FRAENKEL-CONRAT, H. (1966). *Biochemistry* **5**, 2735-2743.

SUN, L., and SINGER, B. (1974). *Biochemistry* **13**, 1905-1912.

SWANN, P. F., and MAGEE, P. N. (1968). *Biochem. J.* **110**, 39-47.

SWANN, P. F., and MAGEE, P. N. (1971). *Biochem. J.* **125**, 841-847.

TSUGITA, A. (1962). *J. Mol. Biol.* **5**, 293-300.

TSUGITA, A., and FRAENKEL-CONRAT, H. (1962). *J. Mol. Biol.* **4**, 73-82.

NUCLEAR INSTRUMENTS AND METHODS 116 (1974) 21–24; © NORTH-HOLLAND PUBLISHING CO.

# A FACILITY FOR CALIBRATIONS AND SHORT EXPERIMENTS WITH HEAVY IONS AT THE BEVATRON

DOUGLAS E. GREINER, PETER J. LINDSTROM, FRED S. BIESER

*Space Sciences Laboratory, University of California, Berkeley, California 94720, U.S.A.*

and

H. H. HECKMAN

*Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, U.S.A.*

Received 23 October 1973

Heavy-ion beams at the Lawrence Berkeley Laboratory Bevatron are routinely available for calibration purposes. These beams are very useful to cosmic-ray physicists and others who wish to have empirical evidence indicating their instruments *really* can identify elements or isotopes. Virtually all isotopes through $^{16}O$ can be produced over a wide energy range.

## 1. Introduction

Since the initial acceleration of heavy ions at the Bevatron[1]) there has been an increasing number of short runs with heavy ions to calibrate cosmic-ray instruments. Many of these experiments have used the magnetic spectrometer system set up by the Heckman group for measuring fragmentation cross sections (Bevatron experiment 162-H). The purpose of this short paper is to gather together the properties of this spectrometer system and some facts about heavy-ion beams at the Bevatron in order to make this information more readily available to physicists who wish to calibrate an instrument or have some other short experiment in mind (see tables 1 and 2). The information presented here should make it possible to write proposals for this class of experiments without reference to the complete Bevatron Experiments' Handbook (UCRL 17333). However, the Handbook is a necessity for proposing major experiments.

TABLE 1

Currently available beams.

| | |
|---|---|
| Particle | $^{2}H$, $^{4}He$, $^{12}C$, $^{14}N$, $^{16}O$. |
| Energy | 0.2–2.1 GeV/N. Typical extraction energies are 0.28, 1.0 and 2.1 GeV/N. Intermediate energies can be achieved with degrading. |
| Energy spread | 300 keV/N fwhm. |
| Flux | Variable $10^{1}$–$10^{6}$ particles per pulse. Particles per pulse can generally be held within a factor of 2 between successive pulses. |
| Spill | 0.2–1.0 s (shorter spill at the lower energy extraction). |
| Spill structure | Generally no worse than 20 to 1 range of instantaneous rate within a spill. |
| Pulse frequency | 10–17 pulses per minute, the higher rate available at low extraction energy. |
| Emittance | $\epsilon_{vert} \leqslant 26$ cm mrad, $\epsilon_{horiz} \leqslant 5$ cm mrad (90% of beam). |

TABLE 2

Experimental facilities.

| | |
|---|---|
| Experiment space | There is a space of 12 ft. beyond the guide rail with a 12 ft. ceiling in the cave. The beam line is 4 ft. above the floor and smaller experiments can be placed on the guide rail support which is 16″ below beam line. |
| Assistance | The Bevatron assigns a staff member to each experiment to assist and communicate. |
| Signals | Timing signals are available to allow predictions of beam spill, etc. |
| People space | There is an experimental house for visitors located outside the cave (50 ft. cable distance). The floor area is 200 sq. ft. |

## 2. Spectrometer properties

The spectrometer arrangement is shown in fig. 1. This apparatus is located in Channel II of the external beam area of the Bevatron. The basic design requirement was to be able to bring fragments from the target area to a focus at the guide rail, over a large rigidity

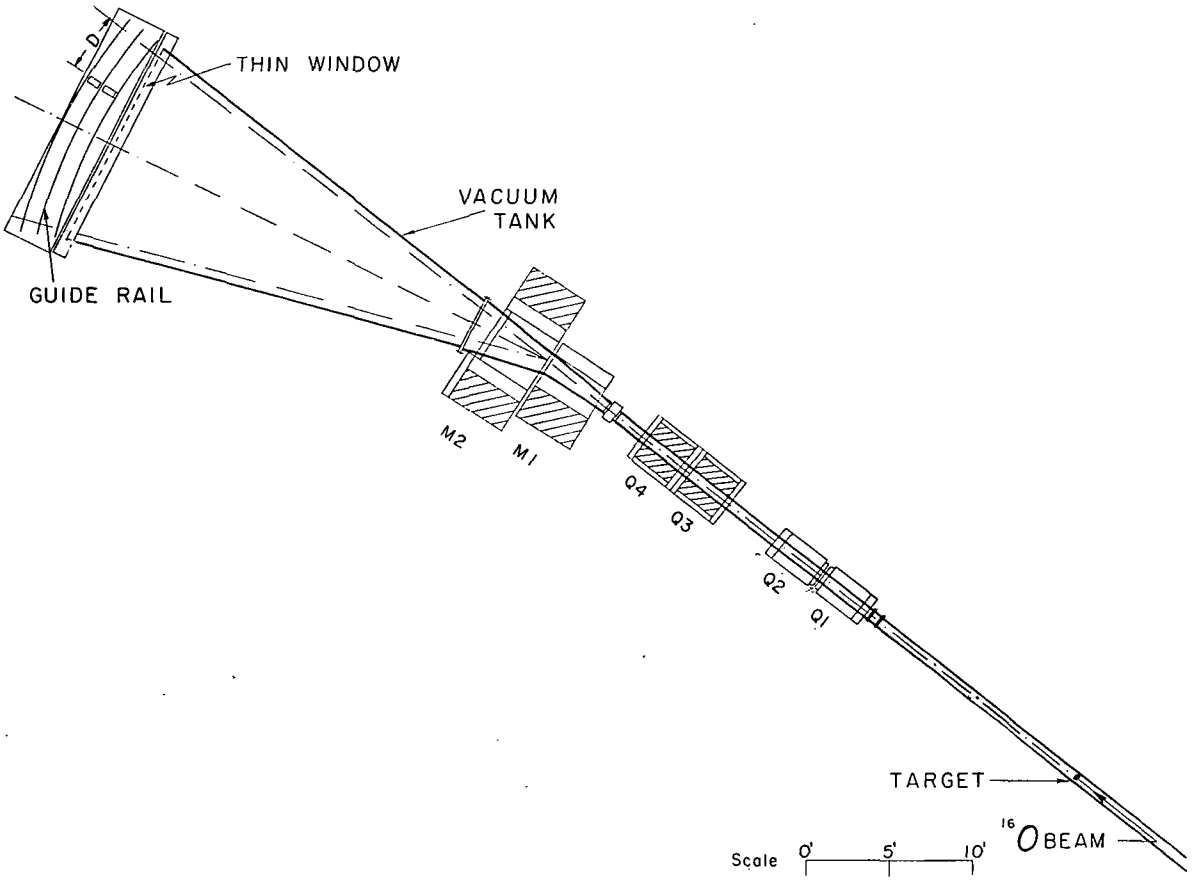



Fig. 1. Focusing magnetic-spectrometer arrangement. Fragments from the target area within .10 mm of 0° are focused at the guide rail. The system can focus rigidities up to 9 GV.

range, while keeping the primary beam in vacuum to reduce background.

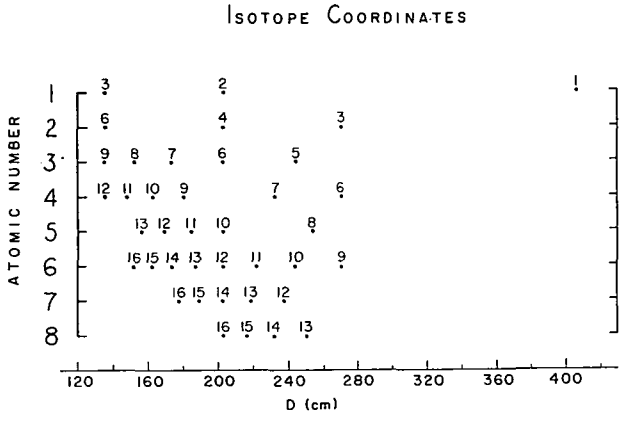



Fig. 2. Plot of isotope deflections for fragmentation on a $^{16}O$ beam centered at $D = 200$. The numbers plotted refer to the mass in amu of the isotope.

The Bevatron's extracted beam (labeled $^{16}O$ beam in fig. 1) enters at target area, which presently is an air gap to allow placement of degraders, scintillators, and targets. The quadrupoles $Q1$–4 focus any rigidity up to 9 GV/$c$ on the guide rail while M1 + M2 produce at the guide rail a deflection $D$ which is inversely proportional to the rigidity. The rail coordinates range from 0.0 cm for an undeflected beam to 400.0 cm at maximum deflection.

The great utility of this arrangement for experiments requiring selected isotopes can be seen as follows: If a primary beam of rigidity $R$ is deflected to appear at $D \approx 200$ cm with a target in place, then fragments of rigidity $0.5R$ to $2R$ can be focused at the guide rail. Because fragments of the projectile move with velocities very close to the initial beam velocity[2]) their rigidity and hence their deflection is determined only by their mass–charge ratio through the relationship.

$$D(Z, A) = 400(Z/A), \qquad (1)$$

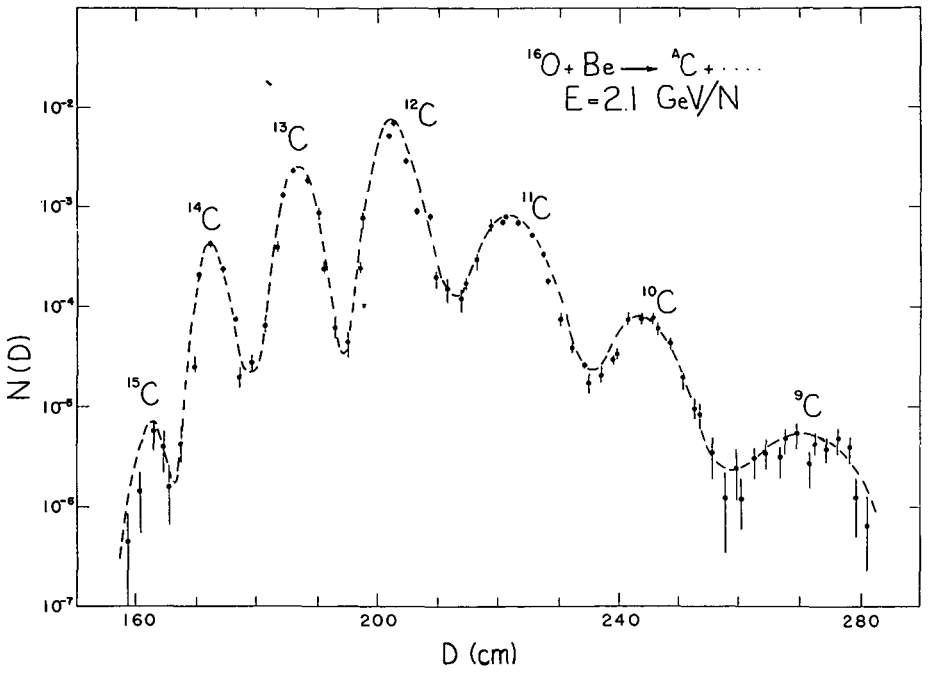



Fig. 3. Plot of carbon isotopes observed from the fragmentation of a $^{16}O$ beam in a Be target. $N(D)$ is the flux integrated over a 1.25 cm$^2$ area at the guide rail per incident beam particle.

where $Z$ and $A$ are, respectively, the charge and mass (amu) of the fragment. In fig. 2, the rail coordinates are shown for all isotopes observed for a primary beam of $^{16}O$. The flux of the heavier isotopes is generally confined to a few centimeters on either side of the central value given by eq. (1). This can be seen for the carbon isotopes produced by a $^{16}O$ beam on a $^{9}Be$ target at 2.1 GeV/N in fig. 3. The ordinate in fig. 3 is the number of particles in an area of approximately 1.25 cm$^2$ at the guide rail per beam particle incident on the $\approx 4$ g target. Thus with $10^6$ beam particles per pulse available, appreciable fluxes of almost any isotope can be produced. The only restriction is that the isotope have a lifetime of the order of the transit time $t$ from the target to the rail. At beam energies of 2.1 GeV/N, the transit time (80 ns) corresponds to a proper time $t = 80/\gamma = 25$ ns.

All of the magnets in the spectrometer are computer controlled and deflections may be read from a scale on the guide rail. A program exists which requires as input the rigidity of the beam, $A$ and $Z$ of the desired isotope, and a rail position, $D$. The computer will then focus the spectrometer magnets so that the isotope appears at $D$.

In summary, this system, with suitable degraders at the target area or at the guide rail, can deliver almost any isotope up to $^{16}O$ at any energy up to 2.1 GeV/N.

## 3. Future developments

In 1974, the transfer line between the Bevatron and the Super Hilac injector (Bevalac) will be operational. The Bevalac will produce more intense heavy-ion fluxes with beams of ions as heavy as $^{56}Fe$. Also in 1974 a 50 MeV linac will be introduced as the primary proton source for the Bevatron. This will allow the present source to be devoted exclusively to heavy ions, thus reducing to a few hours the time required to convert to heavy-ion operation.

## 4. Proposals

Experiment proposals should include:

1) a short summary of the experiment,
2) energies and particles desired,
3) space and power requirements,
4) estimates of set-up and data-taking times,
5) dates when your group can be ready to proceed with the experiment.

All proposals are reviewed by the Bevatron Scheduling Committee which meets formally about every three months. However, short experiments of the type we are discussing here generally fall into a class of experiments which can be treated less formally. We quote from the Bevatron Experimenters' Handbook:

*Another class of experiments is called 'Parasitic'.*

*The general definition of a parasitic experiment is one that requires a negligible amount of Bevatron beam, equipment, and manpower. In addition it must not interfere with the formal Bevatron schedule. If an experiment fits these rather stringent criteria, it can be approved by the ex-officio members of the Committee together with the Bevatron Operations Group without a formal session of the Scheduling Committee.*

Proposals should be sent to

W. D. Hartsough
Building 50, Room 150
Lawrence Berkeley Laboratory
Berkeley, Calif. 94720, U.S.A.
Telephone: 415-843-2740, x5501.

Questions involving scheduling or fees should also be addressed to Mr. Hartsough.

Any questions of a technical nature involving the heavy-ion Bevatron beams and the spectrometer described above can be referred to either

Douglas Greiner or Peter Lindstrom
Building 50, Room 245
Lawrence Berkeley Laboratory
Berkeley, Calif. 94720, U.S.A.
Telephone: 415-843-2740, x5685.

## References

[1]) H. A. Grunder, W. D. Hartsough and E. J. Lofgren, Science **174** (1971) 1128.

[2]) H. H. Heckman, D. E. Greiner, P. J. Lindstrom and F. S. Bieser, Phys. Rev. Lett. **28** (1971) 926.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 249, No. 4, Issue of February 25, pp. 1103-1108, 1974
*Printed in U.S.A.*

# Biosynthesis of Pseudouridine in Transfer Ribonucleic Acid*

(Received for publication, May 14, 1973)

RICCARDO CORTESE,‡ HAROLD O. KAMMEN, SYLVIA J. SPENGLER,§ AND BRUCE N. AMES

*From the Department of Biochemistry and Space Sciences Laboratory, University of California, Berkeley, California 94720*

## SUMMARY

**An enzyme that catalyzes the formation of specific pseudouridine residues in the anticodon region of tRNA has been characterized in extracts of *Salmonella typhimurium* and *Escherichia coli*. It has previously been shown that *hisT* regulatory mutants of *S. typhimurium* lack the pseudouridine modifications in the anticodon region of histidine and several other tRNA species. This has permitted the development of a simple radiochemical assay for the modification enzyme, based on the release of tritium from *hisT* tRNA isolated from cells grown in [5-$^3$H]uridine. This enzyme activity is absent from *hisT* mutants and is relatively heat-labile in extracts from temperature-sensitive *hisT* mutants, which suggests that the modification enzyme itself is the primary gene product of the *hisT* gene.**

Among the various minor nucleosides present in tRNA, pseudouridine is the most common. All known tRNA sequences contain a pseudouridine ($\psi$) residue in the "T$\psi$CPu" side-arm loop (1) except for yeast $tRNA_F^{Met}$ (2) and a $tRNA^{Gly}$ not used in protein synthesis (3). Pseudouridine residues have been located in other regions of the primary structure of tRNA. In many tRNA species from *Escherichia coli* and *Salmonella typhimurium* 1 or 2 pseudouridine residues have been found in the stem and loop of the anticodon region (4–10).

During the study of the regulation of histidine biosynthesis in *S. typhimurium*, we discovered that *hisT* mutants (11), one of the six classes of mutants derepressed for the histidine operon (12), contained a $tRNA^{His}$ lacking 2 pseudouridine residues which are normally present in the anticodon region of wild type $tRNA^{His}$ (13). In contrast, the pseudouridine in the T$\psi$CPu loop was present in both the wild type and mutant $tRNA^{His}$ (12). We have also presented evidence that, as a consequence of the *hisT* mutation, the chromatographic properties of several other species of tRNA are altered without greatly affecting the function of these tRNAs in protein synthesis (13).

* This work was supported by United States Public Health Service Grants AM12092 (B. N. A.) and AM11823 (H. O. K.).

‡ On leave of absence from La Cattedra di Chimica Biologica, 2nd Medical School, University of Naples, Italy.

§ Supported by United States Public Health Service Training Grant 5-TO1-GM-00829-11.

Such findings strongly implied that the *hisT* gene product might be an enzyme that catalyzes the formation of pseudouridine in the anticodon region of specific tRNA species. If this were true, then bulk tRNA from a *hisT* mutant should contain a significant number of residues specifically capable of modification to pseudouridine by enzymes from wild type strains. This situation is analogous to the ability of wild type extracts to modify the submethylated species of tRNA that are synthesized by mutants with defective methylation enzymes (14), or the submethylated and subthiolated tRNA produced by relaxed amino acid auxotrophs (15, 16).

In this paper, we identify an enzymatic activity in wild type strains of *S. typhimurium* and *E. coli* which can modify specific uridine residues of bulk *hisT* tRNA. This activity is not detectable in extracts derived from *hisT* strains and is heat-labile in temperature-sensitive *hisT* strains. These findings offer further evidence that this pseudouridylation enzyme represents the *hisT* gene product.

## MATERIALS AND METHODS

*Chemicals*—Chemicals used for the growth of the cells and for the extraction and analysis of nucleic acid were of reagent grade or otherwise of the highest available purity. [5-$^3$H]Uridine (20,000 mCi per mmole), [6-$^3$H]uridine (20,000 mCi per mmole), and $^{32}P_i$ (carrier-free) were purchased from Schwarz-Mann. DEAE-cellulose paper (Machrey, Nagel and Co.), T1 ribonuclease (Calbiochem), and bovine pancreatic RNase A (Worthington) were purchased from commercial sources.

*Bacterial Strains*—Isogenic *hisT* (TA253) and wild type *S. typhimurium* LT-2 (TA265) were constructed by George Chang (Univ. of California, Berkeley) by transducing *hisT1504* and *hisT*$^+$ into an *aroD purF* double auxotroph (11).

*Growth of Cells and Preparation of Extracts*—Cells were grown in Vogel-Bonner minimal E medium with 0.2% glucose (17) at 37°, harvested in late log phase ($A$ = 1.0 at 650 nm), and kept frozen for several months. Cells were suspended in 2 volumes of buffer containing 10 mM Tris-HCl buffer, pH 7.8, 1 mM EDTA, 1 mM 2-mercaptoethanol, 10 mM $MgCl_2$, and 10% glycerol (v/v), and were sonicated three times for 30 s each. After an initial centrifugation at 30,000 × $g$ for 30 min, the supernatant was respun at 300,000 × $g$ for 90 min. Such extracts were used directly for all of the work reported here.

*Preparation of $^3$H-labeled tRNA*—Cells were grown in 500 ml of the minimal glucose medium (17) containing 10 mCi of [5-$^3$H]-uridine. When they reached the late log phase, they were centrifuged at low speed, suspended in 4 ml of 10 mM Tris-HCl, pH

7.5, and 10 mM $MgCl_2$ (Buffer A), disrupted by sonication, then centrifuged at 30,000 × *g* for 30 min. The supernatant was respun at 100,000 × *g* in a Spinco SW 39 rotor for 4 hr. The supernatant was extracted with distilled phenol saturated with Buffer A, and the tRNA was further purified as previously described (18). The specific activity averaged 10 μCi per $A_{260}$ unit of tRNA. Of the incorporated radioactivity, 63% was found in CMP and 37% in UMP and UMP derivatives, on the basis of alkaline and enzymatic hydrolysis of the product. Both methods of hydrolysis gave identical results.

*Preparation of [$^{32}$P]tRNA and Nucleotide Analysis*—Preparation of [$^{32}$P]tRNA was done according to the procedure described by Singer and Smith (10). After enzymatic modification of the tRNA (see below), the reaction was stopped by the addition of 1 volume of twice distilled phenol, neutralized with Buffer A. Carrier yeast tRNA (50 μg) was added and the samples were shaken for 1 hr at 0°. Approximately 80% of the aqueous phase was carefully removed with a micropipette and the operation was repeated after addition of 0.1 ml of water to the remaining sample. The aqueous phases were combined and NaCl was added (final concentration, 0.5 M), followed by 2.5 volumes of ethanol. The samples were left at −20° overnight and were centrifuged in a clinical centrifuge. The small pellets were resuspended in 10 μl of water and subjected to alkaline hydrolysis, as described earlier (10). The hydrolysates were resolved initially by electrophoresis on DEAE-cellulose paper in pyridine-acetate buffer, pH 3.5, for 3 hours at 2500 volts. UMP and ψMP migrate together in this system. After radioautography, the spot containing the two nucleotides was eluted, taken to dryness and chromatographed on Whatman No. 1 paper, using the solvent, isopropyl alcohol-HCl-water (75:17:1, v/v/v). The UMP and ψMP spots were well separated in this system (19) and after radioautographic identification, they were cut from the paper and counted by Cerenkov radiation in a Nuclear-Chicago Mark II scintillation spectrometer.

*Enzyme Assays*—The tritium release assay for pseudouridylation activity was carried out as follows. The incubation mixtures, 0.1-ml total volume, contained 100 mM glycine-NaOH buffer, pH 9.0; 20 mM $MgCl_2$; 0.02 $A_{260}$ unit of tritiated tRNA; and enzyme. The reaction was started by addition of enzyme, and the samples were incubated at 37°. The reaction was terminated by addition of 1 ml of 12% activated charcoal (Norit A) in 0.1 N HCl (w/v). The samples were left at room temperature for 20 min, with occasional stirring, and then were filtered onto membrane filters (0.45-μ pore size). The filtrate was collected directly into scintillation vials and, after addition of 10 ml of Bray's solution, was counted in a scintillation spectrometer with an over-all efficiency of 15%. Control incubation mixtures from which the enzyme had been omitted (Table I) or which were sampled at zero time gave blanks in the range of 40 to 60 cpm. The data in most of the figures shown have been corrected for these blank values.

Pseudouridylate synthetase was assayed according to Breitman (20).

## RESULTS

*Identification and Specificity of Enzyme Activity*—During the synthesis of pseudouridine, the formation of the C—C bond between $C_5$ of the pyrimidine ring and $C_1$ of the ribose must be accompanied by the release of a proton into the medium (Fig. 1). The availability of pyrimidine derivatives, tritiated at $C_5$ of the ring, has permitted the development of highly sensitive tritium release assays for a number of enzyme reactions involving a similar proton release (20–24). We have used such a tritium release procedure, starting with a labeled tRNA substrate from a *hisT* mutant for the detection of the pseudouridylation activity.

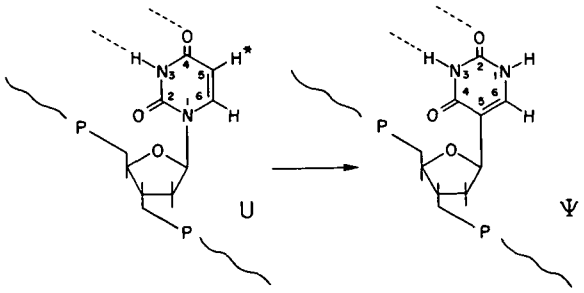



FIG. 1. Tritium release during formation of tRNA pseudouridine residues. The scheme illustrates a possible mechanism for pseudouridine formation. As shown, the generation of a bond between $C_5$ of the pyrimidine ring and $C_1$ of the pentose necessarily involves loss of the hydrogen from the $C_5$ of the ring.

TABLE I

*Release of $^3$H in pseudouridylation assay*

The incubation mixture contained 40 pmoles of tritiated tRNA (1200 cpm per pmole) from the wild type strain or mutant *hisT1504;* 100 mM glycine-NaOH buffer, pH 9.0; 20 mM $MgCl_2$; and 200 μg of enzyme protein in a final volume of 0.1 ml. The reaction was carried out for 30 min at 37° and terminated, and the activity was measured as described under "Materials and Methods." The counts per min released are for the entire reaction mixture.

| Enzyme source | Wild-type tRNA: Experiment 1 | *HisT* tRNA | |
|---|---|---|---|
| | | Experiment 1 | Experiment 2 |
| | *cpm* | *cpm* | |
| No enzyme | 47 | 51 | 22 |
| *HisT1504* | 45 | 56 | 31 |
| *HisT1309* | 51 | 54 | |
| Wild type | 57 | 435 | 573 |
| Wild type (zero time) | | | 34 |
| *Escherichia coli* B | | | 560 |

The tRNA, tritiated at $C_5$ of the pyrimidine ring, was prepared from a *hisT* mutant of *S. typhimurium* grown in the presence of [5-$^3$H]uridine (see "Materials and Methods"). In the presence of a pseudouridylation activity, tritium should be released into the medium at a rate proportional to the enzyme activity. The data shown in Table I demonstrate the presence of such an activity in extracts from wild type *S. typhimurium.* Under the same conditions, no tritium release occurred with labeled wild type tRNA, which indicates that this tRNA lacked modifiable pyrimidine residues. It was also possible to detect this activity in extracts from *E. coli* B which had about the same activity as found in *S. typhimurium.*

The tritium release activity was also tested with extracts from two independently isolated *hisT* mutants: *hisT1504*, the source organism for the labeled tRNA and the strain used for determination of the $tRNA^{His}$ sequence (13); and *hisT1309*, an amber mutant in the same gene (11). Tritium release activity was undetectable in extracts from both *hisT* strains.

With *hisT* tRNA labeled at $C_5$, the rate of tritium release was linear for 5 or 10 min and continued for 30 min or more under the conditions shown in Fig. 2. On the other hand, tRNA from the same mutant, but tritiated at the $C_6$ position of the pyrimidine residues, did not release any tritium, even after extended

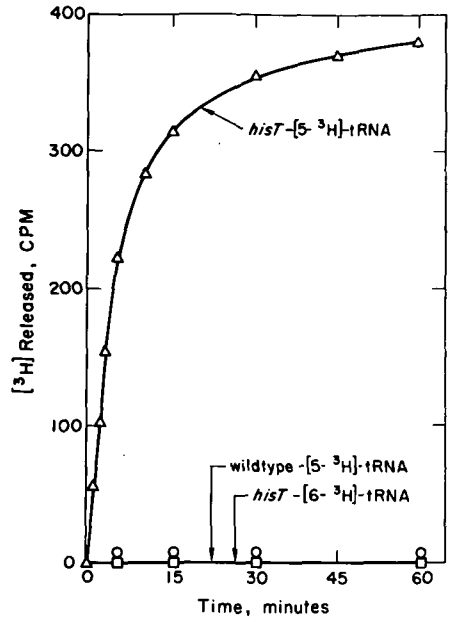


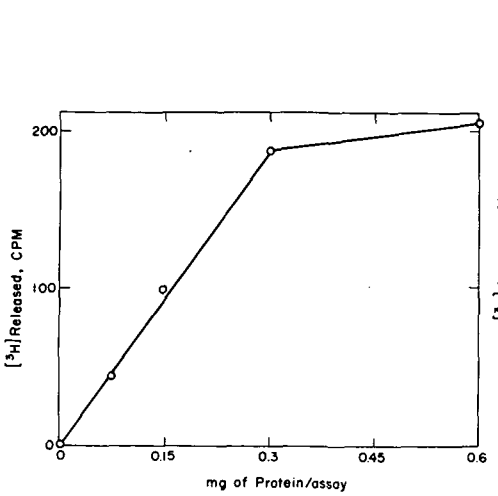

FIG. 2. Rate and specificity of tritium release. Assay conditions were identical with those of Table I, except that 100 $\mu$g of the protein were used.

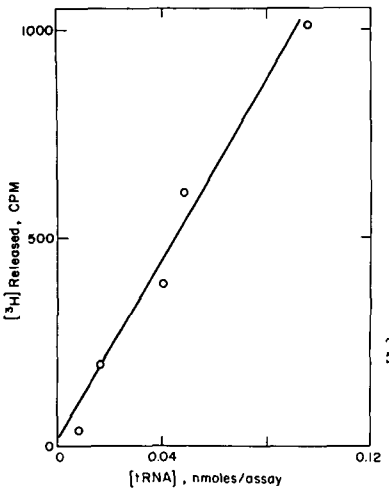
reaction of about 0.4 mole of uridine per mole of tRNA.[1] This value may not necessarily represent the maximum number of modifiable sites in the tRNA, since the crude extracts used in the experiments might have contained inhibitors, nucleases, or other contaminants affecting the plateau. This aspect of the reaction deserves further examination with purified enzymes.

In Fig. 5, the initial rates of tritium release are plotted as a function of tRNA concentration. From the Lineweaver-Burk plot (see *inset*), an apparent $K_m$ of approximately $10^{-6}$ M is obtained. This value is a maximum estimate, since it is based on the assumption that every tRNA molecule is a substrate for the enzyme. However, since not all molecules of tRNA contain pseudouridine modifications in the anticodon region (1), the affinity of the enzyme for the specific tRNA substrate is probably even greater than the apparent $K_m$ would indicate.

At high concentrations of tRNA, there appears to be a small divergence from ideal Lineweaver-Burk kinetics (see *inset*, Fig. 5). This could indicate that the substrate tRNA mixture contained nonmodifiable sequences or configurations which competed for the enzyme site or which interfered with the enzyme activity. This feature deserves further examination.

Other reaction requirements are given in Figs. 6 and 7. The pH dependence of the reaction (Fig. 6) showed a sharp optimum at 9.0. The reaction also showed an absolute requirement for



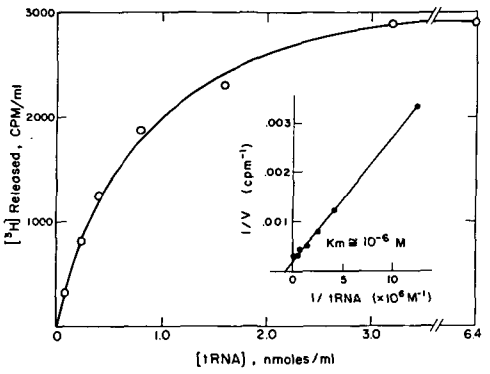




FIG. 3 (*left*). Rate of tritium release as a function of enzyme concentration. Reaction conditions were those shown in Table I. The incubation was carried out for 3 min at 37°.

FIG. 4 (*center*). Relationship between the plateau of tritium release and tRNA concentration. The assay conditions were those given in Fig. 2. The incubations were carried out for 60 min in order to achieve a plateau of tritium release (see Fig. 2).

FIG. 5 (*right*). Initial rate of tritium release as a function of *hisT* tRNA concentration. The reaction conditions were those of Fig. 2. The incubation time was 3 min. The $V$ values shown in the *inset* refer to counts per min of $^3$H released.

incubation (Fig. 2). This provides further support for the specificity of the reaction and suggests that the mechanism does not involve any randomization of hydrogen atoms at carbon 6 of the pyrimidine ring.

*Kinetic Features of the Enzyme Reaction*—Fig. 3 depicts the tritium release activity as a function of the enzyme concentration. With *hisT* tRNA at 0.4 nmole per ml, the rate of tritium release was a linear function of the concentration of enzyme, up to a saturating level of 3 mg per ml of protein.

The maximum extent of tritium release was determined by prolonged incubation with subsaturating levels of enzyme and was directly proportional to the initial concentration of tRNA (Fig. 4). From the specific activity of the tRNA and its uridine content, we estimate that the plateau levels correspond to the

$Mg^{2+}$ (data not shown) and was optimal above a $Mg^{2+}$ concentration of 1.0 mM.

The activity was highly sensitive to increased salt concentration (Fig. 7). At 65 mM NaCl, the activity was only 50% of the optimum rate. However, full enzyme activity was maintained

[1] This value has been determined as follows. The concentration of tRNA was calculated from an assumed absorbance of 20 $A_{260}$ per mg, and an average molecular weight of 25,000. Accordingly, 0.02 $A_{260}$ unit was equivalent to 1 $\mu$g or 40 pmoles of tRNA. Each picomole of *hisT* tRNA was assumed to contain 14 pmoles of uridine (25), which accounted for 37% of the $^3$H, as determined by direct experiment (see "Materials and Methods"). With a specific activity of 1,200 cpm per pmole of tRNA, the modification of 1 pmole of uridine would correspond to the release of 1,200 × 0.37/14 = 31.7 cpm of $^3$H.

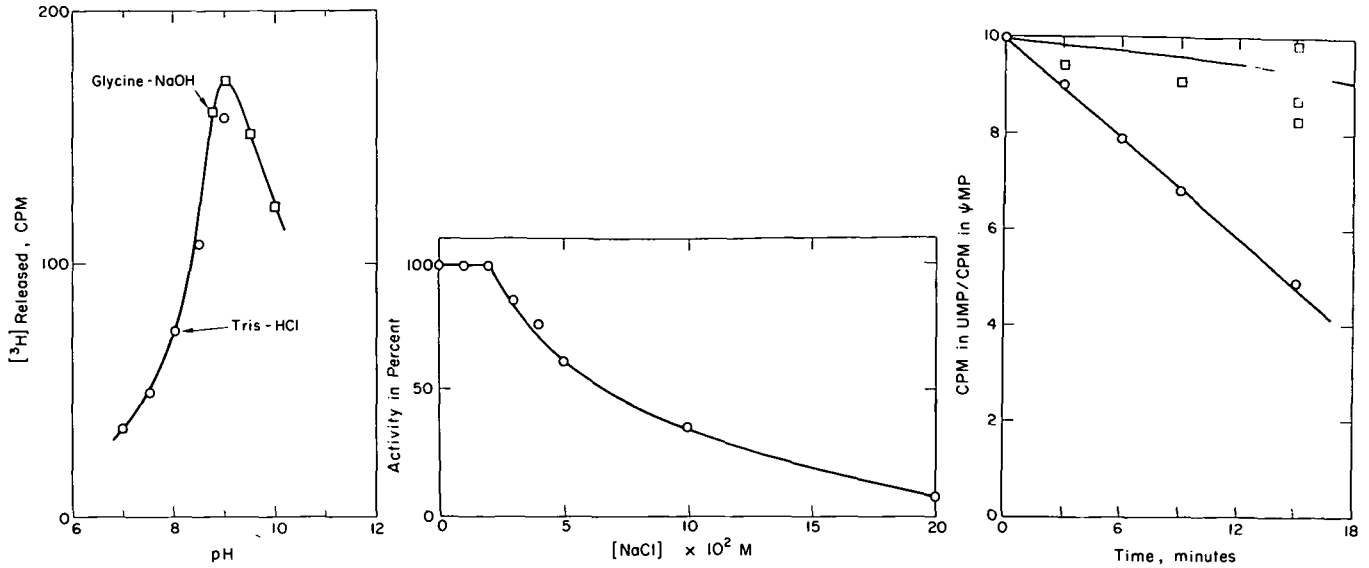



FIG. 6 (*left*). pH Dependence of tritium release. The standard assay procedure was used, but the buffer pH was varied as shown, at a concentration of 100 mM. The assay mixtures contained 20 μg of enzyme protein and were incubated for 5 min.

FIG. 7 (*center*). Effect of salt concentration on rate of tritium release. The reaction conditions were those given in Table I, except that 20 μg of enzyme protein were used, and NaCl was added to the final concentrations shown. The incubation was for 5 min at 37°.

FIG. 8 (*right*). Formation of tRNA pseudouridylate during incubation of *hisT* tRNA with enzyme extracts. The incubation mixtures contained 100 mM glycine buffer, pH 9.0; 20 mM $MgCl_2$; 100 pmoles per ml of $^{32}P$ - labeled *hisT* tRNA and 500 μg per ml of enzyme protein from either the wild type strain or from strain *hisT1504*. At the designated time points, 0.1-ml samples were removed and mixed with an equal volume of neutralized phenol. Carrier yeast tRNA was added and the nucleic acids were isolated, hydrolyzed, and subjected to nucleotide analysis as described under "Materials and Methods."

when the salt concentration was restricted to 20 mM or less. The inhibition by salts appeared to be nonspecific, since $(NH_4)_2SO_4$, $NH_4Cl$, KCl, and NaCl were equally inhibitory at a concentration of 100 mM.

The extract was readily desalted by passage through Sephadex G-50, but the activity was sometimes lost after dialysis. The preparation could be stored for at least 10 days in 10% glycerol with no loss of activity, but freezing and thawing led to severe losses.

*Characterization of Pseudouridylate Formation by Enzyme Extracts*—In order to verify the occurrence of the pseudouridylate modification, it was necessary to demonstrate that the action of the enzyme led to the disappearance of uridylate or cytidylate residues, with the concurrent appearance of pseudouridylate. This was carried out by incubation of [$^{32}P$]tRNA from the *hisT* mutant with wild type enzyme extract. As a control, the *hisT* tRNA was treated with homologous enzyme from the same *hisT* strain. The results are shown in Fig. 8. In order to normalize the results for variable losses during the analytical steps, the activity is expressed in terms of the ratios of UMP:ψMP during the course of the reaction. For bulk *hisT* tRNA, this ratio is equal to 10, whereas it is equal to 5 for wild type tRNA. When the *hisT* tRNA was treated with enzyme from wild type cells, this ratio decreased to 5 within a 15-min incubation period (Fig. 8). In the control incubation, the ratio changed only slightly, from 10 to approximately 9, during the same time course.

It is doubtful whether the small change in UMP:ψMP ratio in the control incubation has any significance, particularly since such homologous reaction mixtures were inactive in the tritium release assay (Table I). In other experiments (not shown) we have found no change in the ψMP content of *hisT* tRNA after incubation with homologous extracts. We suspect that the changes in the control incubation were a result of the selective cleavage of UMP residues by the mixture of nuclease activities in the crude cell extracts.

*Defective Enzyme Activity in a Temperature-sensitive Mutant*—Further evidence that *hisT* mutants are altered in a protein responsible for the formation of pseudouridine residues has been obtained from the study of a temperature-sensitive *hisT* mutant. This mutant, TA263, is subject to normal regulation of the histidine operon when it is grown at 30°, but is derepressed 2- to 3-fold at 37° and 8-fold at 42° (11). If the *hisT* gene product is an enzyme that catalyzes formation of tRNA pseudouridine residues, then the enzyme activity itself should be temperature-sensitive in extracts from strain TA263.

This has proven to be the case when the activity is measured by the tritium release assay (Fig. 9). When the extracts were heated at 60°, the activity from strain TA263 was inactivated 6 to 7 times as rapidly as the activity in wild type extracts.

## DISCUSSION

Despite a variety of early conflicting evidence (reviewed in Ref. 26), it is now abundantly clear that the pseudouridine modifications in tRNA occur after assembly of the tRNA chain (27, 28), as is the case with all other known tRNA modification reactions.

The tRNA species which accumulate in *hisT* mutants of *S. typhimurium* provide a natural source of unmodified loci for a specific group of pseudouridine modifications. By the use of *hisT* tRNA, tritiated specifically at the $C_5$ pyrimidine residues, the tritium release assay provides a convenient and rapid measure of at least a partial reaction involved in pseudouridine formation. The specificity of this assay is implicit from the following considerations. (*a*) Tritium release occurs only when the tRNA from *hisT* cells is treated with wild type extracts (Fig. 2). No release is found when wild type tRNA is used, since all potential sites are already modified. No release is obtained with extracts

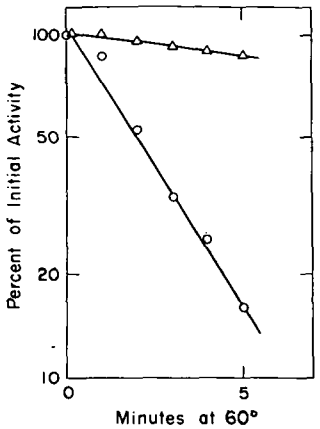



FIG. 9. Temperature sensitivity of wild type and *hisT* extracts. Tubes containing 30 μg of enzyme protein in 100 μl of cell extract were warmed in a 60° bath for the intervals shown and then were immediately chilled in an ice bath. The enzymatic activity was then measured by the tritium release assay, as described under "Materials and Methods." The activity of the untreated extract from the *hisT* mutant, TA263, was only 60% of the activity of the untreated wild type extract. The values shown are normalized to 100% activity for each untreated extract.

from *hisT* cells, since they lack the enzyme. (*b*) The replacement of pseudouridine residues by uridine residues in the anticodon region of $tRNA^{His}$ (13) and $tRNA^{Leu}$ (29) of *hisT* mutants has been established by sequencing. In addition, the chromatographic properties of other species of tRNA are altered in *hisT* mutants, notably those of $tRNA^{Tyr}$ (13) and $tRNA^{Lys}$ (Footnote 2) and the major (13) and minor species of $tRNA^{Leu}$.[3] This pattern is consistent with the known presence of pseudouridine in this position in these tRNA species. (*c*) Although other modification reactions would be measured by the tritium release assay, *e.g.* the formation of 5-methylated or other 5-substituted pyrimidine derivatives (30), it is unlikely that such reactions contribute to the tritium release to any significant extent. This conclusion is based on several facts. First, no additional cofactors were necessary for tritium release, as might be expected for other modification reactions. Second, release of tritium by wild type extracts took place only with *hisT* tRNA as substrate; wild type tRNA, labeled in the same manner, did not release tritium upon treatment with wild type extract. Finally, there is no evidence that *hisT* tRNA has defects in modification other than the absence of certain pseudouridine residues (13).

This assay system may also be suitable for characterizing the pseudouridine modification enzymes of other organisms, particularly by employing heterologous reactions, in which the labeled tRNA is derived from organisms of low pseudouridine content. Such heterologous reactions are readily demonstrable with the tRNA methylation enzymes (31, 32).

Several insights into the enzyme mechanism can be derived, even with the limited information available at this time. (*a*) The enzyme described here is not responsible for the pseudouridine modification in all regions of the tRNA molecule. This is evident from the fact that *hisT* mutants, which lack this activity, still produce tRNA with pseudouridine in the TψCPu loop (13). This residue therefore must be modified independently, presumably by a second enzyme which may act on a tRNA precursor species (33). It is not known whether the pseudouridine residues of ribosomal RNA (34, 35) are formed by still another enzyme or not. For these reasons, we propose the designation tRNA pseudouridine synthetase I for the enzyme described here, which is absent or defective in *hisT* mutants, and tRNA pseudouridine synthetase II for the activity implicated in the TψCPu sequence. (*b*) Although there has been some speculation that the formation of pseudouridine could derive from a cytidine residue (see Ref. 24, p. 310), the most likely precursor would appear to be an unmodified uridine residue. This likelihood is implicit in the normal base-pairing relationships of the pseudouridine residues in the anticodon stem (10, 13), and in the fact that *hisT* tRNA contains uridine, rather than cytidine residues, in place of the missing pseudouridine residues of wild type tRNA (13). (*c*) The reaction mechanism probably does not involve loss of a hydrogen atom from carbon 6, since *hisT* tRNA tritiated at carbon 6 does not react with wild type enzyme (Fig. 2).

Breitman (20) has used this type of tritium release assay to characterize the enzyme, pseudouridylate synthetase, which catalyzes the formation of ψMP from uracil and ribose-5-P, in extracts from various strains of *E. coli*. We have been unable to detect this activity in our wild type strain of *S. typhimurium* (LT-2) by direct assay. In addition, strain LT-2 is unable to grow on pseudouridine as a carbon source, nor can the growth of *pyr* auxotrophs of LT-2 be supported by pseudouridine. These findings, and the lack of a specific tRNA requirement in the pseudouridylate synthetase reaction, clearly show that it is a distinct enzyme from the tRNA pseudouridine synthetase described here.

The biological effects of the *hisT* mutation are uniquely pleiotropic. The lack of the pseudouridine residues in the anticodon region of $tRNA^{His}$ in some manner affects the production of the histidine biosynthetic enzymes. The same *hisT* mutant is also altered in the regulation of the *leu* and *ile* operons.[3] *HisT* mutants also possess an altered $tRNA^{Lys}$ and are resistant to inhibition by the lysine analog, 2-*S*-aminoethylcysteine.[2] The manner in which the lack of pseudouridine modifications can lead to these regulatory defects remains unclear, but is obviously of great interest.

Finally, the analysis of a temperature-sensitive *hisT* mutant suggests that the *hisT* gene codes for the modification enzyme rather than for a protein which regulates the activity or synthesis of the modification enzyme.

[2] R. Cortese, R. Landsberg, and B. N. Ames, unpublished observations.

[3] R. Cortese, R. Von der Haar, R. Landsberg, H. E. Umbarger, and B. N. Ames, manuscript in preparation.

## REFERENCES

1. STAEHELIN, M. (1971) *Experientia* **21**, 1–11
2. SIMSEK, M., AND RAJBHANDARY, U. L. (1972) *Biochem. Biophys. Res. Commun.* **49**, 508–515
3. ROBERTS, R. J. (1972) *Nature New Biol.* **237**, 44–45
4. FOLK, W. R., AND YANIV, M. (1972) *Nature New Biol.* **237**, 165–166
5. BARRELL, B. G., AND SANGER, F. (1969) *Fed. Eur. Biochem. Soc. Lett.* **3**, 275–278
6. CORY, S., AND MARCKER, K. A. (1970) *Eur. J. Biochem.* **12**, 177–194
7. GOODMAN, H. M., ABELSON, J. N., LANDY, A., ZADRAZIL, S., AND SMITH, J. D. (1970) *Eur. J. Biochem.* **13**, 461–483
8. BLANK, H., AND SÖLL, D. (1971) *Biochem. Biophys. Res. Commun.* **43**, 1192–1194
9. DUBE, S. K., MARCKER, K. A., AND YUDELEVICH, A. (1970) *Fed. Eur. Biochem. Soc. Lett.* **9**, 168–170
10. SINGER, C. E., AND SMITH, G. R. (1972) *J. Biol. Chem.* **247**, 2978–3000
11. CHANG, G. W., ROTH, J. R., AND AMES, B. N. (1971) *J. Bacteriol.* **108**, 410–414

12. Brenner, M., and Ames, B. N. (1971) in *Metabolic Pathways V. Metabolic Regulation* (Vogel, H. J., ed), pp. 349–387, Academic Press, New York
13. Singer, C., Smith, G., Cortese, R., and Ames, B. N. (1972) *Nature New Biol.* **238**, 72–74
14. Björk, G., and Isaksson, L. A. (1970) *J. Mol. Biol.* **51**, 83–100
15. Fleissner, E., and Borek, E. (1962) *Proc. Nat. Acad. Sci. U. S. A.* **48**, 1199–1203
16. Harris, C. L., and Tichener, E. B. (1972) *Biochemistry* **10**, 4207–4212
17. Vogel, H. J., and Bonner, D. M. (1956) *J. Biol. Chem.* **218**, 97–106
18. Brenner, M., and Ames, B. N. (1972) *J. Biol. Chem.* **247**, 1080–1088
19. Barrell, G. B. (1971) in *Procedures in Nucleic Acid Research* (Cantoni, G. L., and Davies, D. R., eds), Vol. 2, p. 768, Harper & Row, New York
20. Breitman, T. R. (1970) *J. Bacteriol.* **103**, 263–264
21. Lomax, M. I. S., and Greenberg, G. R. (1967) *J. Biol. Chem.* **242**, 109–113
22. Kammen, H. O. (1966) *Anal. Biochem.* **17**, 553–556
23. Roberts, D. W. (1966) *Biochemistry* **5**, 3546–3548
24. Yeh, Y-C., and Greenberg, G. R. (1967) *J. Biol. Chem.* **242**, 1307–1313
25. Brown, G. L. (1963) *Progress in Nucleic Acid Research and Molecular Biology*, Vol. 3, pp. 259–310, Academic Press, New York
26. Goldwasser, E., and Heinrikson, R. L. (1966) *Progress in Nucleic Acid Research and Molecular Biology*, Vol. 5, p. 104, Academic Press, New York
27. Johnson, L., and Söll, D. (1970) *Proc. Nat. Acad. Sci. U. S. A.* **67**, 943–950
28. Weiss, S. B., and Legault-Démare, J. (1965) *Science* **149**, 429–431
29. Allaudeen, H. S., Yang, S. K., and Söll, D. (1972) *Fed. Eur. Biochem. Soc. Lett.* **28**, 205–208
30. Hall, R., (1971) *The Modified Nucleosides in Nucleic Acid*, p. 258, Columbia Univ. Press, New York
31. Gold, M., Hurwitz, J., and Anders, M. (1963) *Proc. Nat. Acad. Sci. U. S. A.* **50**, 164–169
32. Baguley, B. C., and Staehelin, M. (1968) *Biochemistry* **7**, 45–50
33. Altman, S., and Smith, J. D. (1971) *Nature New Biol.* **233**, 35–39
34. Fellner, P., and Sanger, F. (1968) *Nature* **219**, 236–238
35. Dubin, D. T., and Günlap, A. (1967) *Biochim. Biophys. Acta* **134**, 106–123

J. Mol. Evol. 3, 189–208 (1974)
© by Springer-Verlag 1974

A-74-39744

# Conservation of Shannon's Redundancy for Proteins

Lila L. Gatlin

Space Sciences Laboratory, University of California, Berkeley, California 94720

Received January 25, 1974

*Summary.* The fundamental question of the randomness or non-randomness of protein sequences is studied utilizing information theory. Information parameters for random protein sequences are derived utilizing Monte Carlo methods. Real sequence parameters are compared with the random standards for sequences of a specific length.

Non-periodic proteins as a whole have random Shannon redundancies ($R$) except for the constraints imposed by their short length and the genetic code. This random $R$ value is conserved within the basic informational unit of proteins throughout the evolutionary process but not the components of $R$, $D_1$, and $D_2$. Highly periodic proteins are identifiable by their high $R$ value and their high $D_1$-low $D_2$ pattern. A parameter for calculating the degree of periodicity is defined. There is an unexpected suggestion of generalized, residual periodicity in proteins as a whole.

*Key words*: Protein Redundancy — Periodicity — Conservation Law.

When we consider the wide variety of proteins found in nature we are struck by the apparent lack of any universal restrictions on their amino acid sequences. Protein sequences which are almost totally unrelated may sometimes perform, to a reasonable degree, the same biological function; and very similar protein sequences may have completely different biological functions.

When amino acid frequencies are averaged over a broad spectrum of non-homologous proteins, this distribution can be predicted *approximately* from random models of DNA involving the genetic code (King and Jukes, 1969; Ohta and Kimura, 1971; Davidson and Dancis, 1972; Jukes and Holmquist, 1972).

The purpose of this work is to examine the randomness or non-randomness of proteins with mathematically precise measures of randomness from information theory. In particular we will utilize two measures of randomness from concepts in information theory which may be calculated from the known amino acid sequence. The first measure, $D_1$, measures roughly the tendency of the chain to use some amino acids more than others, and the second measure, $D_2$, the tendency to use certain amino acid pairs more than others. For a random chain we would expect all singlets and doublets to appear with equal frequency as much as possible for the length of the

sequence. Immediately the length variable becomes a dominant factor in our thinking. We may draw an analogy to a simple game.

If we toss a fair coin $N$ times, we do not necessarily expect $N/2$ heads and $N/2$ tails for each sequence of $N$ trials. However, we can predict the distribution of various types of runs as a function of the number of trials (or length of the sequence).

This is our approach. We will predict the distribution of our measures, $D_1$ and $D_2$, as a function of sequence length for sequences that are known to be randomly generated as in a simple game and then compare this "theoretical" random distribution with the distribution of $D_1$ and $D_2$ values found in nature over a wide variety of non-homologous proteins with a wide variety of lengths. In addition we will also compute the distribution to be expected if the base sequence of DNA were randomly generated and the degeneracy of the genetic code were contributing to $D_1$ and $D_2$.

Shannon's redundancy, $R$, which is proportional to the sum of $D_1$ and $D_2$, will play the role of the invariant. $R$ is conserved in the evolutionary process but not its components, $D_1$ and $D_2$. We may draw a loose, descriptive analogy to a vector whose length is invariant under coordinate transformation, but not its components.

We now proceed to define precisely what we mean by a randomly generated sequence and to define mathematically the information parameters.

## Theory

The basic theory has been presented in book form (Gatlin, 1972) and will be summarized only briefly below.

A *source* is any apparatus which emits a sequence of symbols from a specified alphabet according to specified probabilities. For a Markov source with memory, $m$, the probability of emission of each letter depends on the $m$ symbols preceding it in the sequence. If the Markov source is *ergodic*, the distributions of $n$-tuple probabilities where $n = 1, 2 \ldots (m+1)$ by definition approach stable limiting values for very long sequences which are characteristic of that source.

The total divergence from the maximum entropy state or the "information density", $I_d$, is defined by:

$$I_d^{(m)} = D_1 + D_2 + D_3 + \ldots D_{(m+1)} \tag{1}$$

where

$$D_n = H_M^{(n-2)} - H_M^{(n-1)} \tag{2}$$

where $n$ is the length of the $n$-tuple under consideration and

$$H_M^{(m)} = -\sum_{i_\sigma=1}^{a} p_{i_1} p_{i_1 i_2} p_{i_1 \ldots i_{(m+1)}} \log p_{i_1 \ldots i_{(m+1)}} \tag{3}$$

where $p_{i_1 \ldots i_{(m+1)}}$ is the conditional probability of emission of letter $m+1$, given the preceding $m$ symbols and $a$ is the number of letters in the alphabet. These relations are derived in Gatlin (1972, pp. 62–68). $I_d$ is a measure of the *fidelity* of a message.

For very long sequences, a perfectly random sequence may be defined as one whose informational entropy is maximal. The maximum entropy state for a sequence of symbols is characterized by equiprobable, independent symbols. If the sequence

diverges from the maximum entropy state, the divergence from equiprobability of the symbols, $D_1$, is given by

$$D_1 = H_1^{\text{Max}} - H_1^{\text{Ob}} \tag{4}$$

where $H_1^{\text{Max}} = \log a$ and $H_1^{\text{Obs}}$ is the experimentally observed value. If the sequence of symbols diverges from the maximum entropy state owing only to a divergence from independence of the symbols, this divergence, $D_2$, is given by:

$$D_2 = H_1 - H_M \tag{5}$$

When the superscript $(m)$ is not specifically indicated, the memory of the source is understood to be one. For conceptual purposes it is helpful to consider the special case, $m = 1$. Then

$$I_d = D_1 + D_2 \tag{6}$$

This may also be expressed in terms of the redundancy, $R$, defined by Shannon (1949).

$$R \log a = D_1 + D_2 \tag{7}$$

$R$ is a dimensionless fraction whereas $D_1$ and $D_2$ are expressed in bits.

Often when a biologist uses the phrase "deviation from the random" in describing a sequence of symbols, he means the divergence from independence, $D_2$, of the symbols as separate and distinct from the divergence from equiprobability, $D_1$. But it should be understood that $D_1$ is a "deviation from the random" also because the most random state is the maximum entropy state and this is characterized by events which are both equiprobable and independent.

When a random source emits a very long sequence, $R^{(m)} = 0$ for all $m$ which implies that $D_n = 0$ for all $n$. There is no divergence from equiprobability or independence of the symbols; and the $n$-tuple distributions are all equiprobable for $n = 1, 2 \ldots m + 1$.

A random source is thus, by definition, ergodic because all the $n$-tuple distributions approach these stable, equiprobable states. Beyond $m + 1$ the $E$-property (*cf.* Khinchin, 1957) insures that for ergodic sources all very long sequences of significant probability are all equiprobable also.

However, if we are dealing with short sequences, the sequence is not of sufficient length for all $n$-tuples to be represented equally. Thus $R^{(m)} \neq 0$ for at least some $m$ and a divergence from equiprobability of the $n$-tuples, $D_n$, occurs for some $n$ even though we are dealing with the output of a random source.

Alternatively we may regard the source as random except for the constraint that it emit only short sequences of a given length $L$. Nevertheless, the mean values and their variances for the information parameters as a function of $L$ are quite predictable. The values of the parameters for a given value of $L$ are normally distributed and converge to stable means which may be estimated by Monte Carlo methods. Let us now develop these methods.

## The Random Game

Let us play a simple game of tossing an icosahedral die $L$ times where $L$ is the length of the random protein sequence to be generated. We may simulate this game on a computer as follows.

Most large computing systems contain random number generators as intrinsic functions. These generators typically calculate random numbers from the relation

$$X_{i+1} = A X_i (\text{mod } M) \tag{8}$$

where $A$ is a multiplier and $M$, the modulus, is typically chosen to be $2^k$ for a machine with a $k$-bit accumulator. These generators return pseudo-random numbers uniformly distributed between zero and one.

Imperfections in these generators have been discovered. MacLaren and Marsaglia (1965) pointed out a serial correlation between $X_i$ and $X_{i+1}$ for small enough unit spaces. A periodicity in the trailing bits is also observable.

Table 1. Means and $\sigma$ for random generators

| $L$ and $\sigma$ | Protein | | | 50% (C + G) DNA-to-protein | | | 42% (C + G) DNA-to-protein | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\bar{R}$ | $\bar{D}_1$ | $\bar{D}_2$ | $\bar{R}$ | $\bar{D}_1$ | $\bar{D}_2$ | $\bar{R}$ | $\bar{D}_1$ | $\bar{D}_2$ |
| $L = 10$ | 0.916 | 1.385 | 2.572 | 0.886 | 1.535 | 2.293 | 0.890 | 1.509 | 2.337 |
| $\sigma$ | 0.049 | 0.234 | 0.436 | 0.062 | 0.301 | 0.561 | 0.058 | 0.296 | 0.537 |
| $L = 20$ | 0.825 | 0.797 | 2.771 | 0.801 | 0.938 | 2.523 | 0.799 | 0.953 | 2.501 |
| $\sigma$ | 0.038 | 0.181 | 0.340 | 0.047 | 0.246 | 0.441 | 0.046 | 0.251 | 0.442 |
| $L = 40$ | 0.700 | 0.393 | 2.630 | 0.671 | 0.570 | 2.239 | 0.673 | 0.554 | 2.357 |
| $\sigma$ | 0.024 | 0.107 | 0.201 | 0.031 | 0.160 | 0.283 | 0.030 | 0.171 | 0.287 |
| $L = 60$ | 0.610 | 0.238 | 2.398 | 0.582 | 0.431 | 2.086 | 0.582 | 0.444 | 2.070 |
| $\sigma$ | 0.022 | 0.088 | 0.173 | 0.024 | 0.118 | 0.209 | 0.025 | 0.126 | 0.215 |
| $L = 80$ | 0.536 | 0.197 | 2.120 | 0.519 | 0.374 | 1.867 | 0.517 | 0.380 | 1.853 |
| $\sigma$ | 0.015 | 0.066 | 0.111 | 0.019 | 0.103 | 0.162 | 0.019 | 0.106 | 0.169 |
| $L = 100$ | 0.484 | 0.146 | 1.945 | 0.465 | 0.329 | 1.680 | 0.466 | 0.344 | 1.670 |
| $\sigma$ | 0.014 | 0.044 | 0.084 | 0.015 | 0.086 | 0.128 | 0.017 | 0.078 | 0.124 |
| $L = 120$ | 0.438 | 0.121 | 1.774 | 0.427 | 0.316 | 1.532 | 0.428 | 0.313 | 1.537 |
| $\sigma$ | 0.014 | 0.036 | 0.078 | 0.014 | 0.079 | 0.109 | 0.015 | 0.079 | 0.118 |
| $L = 140$ | 0.402 | 0.093 | 1.644 | 0.392 | 0.292 | 1.401 | 0.394 | 0.304 | 1.397 |
| $\sigma$ | 0.013 | 0.029 | 0.069 | 0.014 | 0.065 | 0.096 | 0.015 | 0.071 | 0.100 |
| $L = 160$ | 0.370 | 0.086 | 1.511 | 0.368 | 0.276 | 1.316 | 0.368 | 0.286 | 1.303 |
| $\sigma$ | 0.011 | 0.030 | 0.056 | 0.014 | 0.065 | 0.081 | 0.015 | 0.069 | 0.098 |
| $L = 180$ | 0.343 | 0.081 | 1.400 | 0.341 | 0.267 | 1.207 | 0.341 | 0.268 | 1.207 |
| $\sigma$ | 0.011 | 0.027 | 0.059 | 0.013 | 0.064 | 0.076 | 0.012 | 0.057 | 0.064 |
| $L = 200$ | 0.320 | 0.069 | 1.314 | 0.324 | 0.261 | 1.137 | 0.324 | 0.265 | 1.134 |
| $\sigma$ | 0.010 | 0.023 | 0.049 | 0.011 | 0.055 | 0.063 | 0.014 | 0.051 | 0.072 |
| $L = 220$ | 0.298 | 0.065 | 1.223 | 0.306 | 0.248 | 1.076 | 0.307 | 0.250 | 1.075 |
| $\sigma$ | 0.012 | 0.018 | 0.053 | 0.013 | 0.049 | 0.064 | 0.012 | 0.054 | 0.069 |
| $L = 240$ | 0.281 | 0.058 | 1.158 | 0.290 | 0.249 | 1.007 | 0.289 | 0.243 | 1.007 |
| $\sigma$ | 0.010 | 0.019 | 0.048 | 0.013 | 0.054 | 0.071 | 0.012 | 0.052 | 0.065 |
| $L = 260$ | 0.261 | 0.055 | 1.072 | 0.275 | 0.234 | 0.956 | 0.276 | 0.244 | 0.949 |
| $\sigma$ | 0.011 | 0.016 | 0.052 | 0.013 | 0.050 | 0.065 | 0.012 | 0.051 | 0.062 |
| $L = 280$ | 0.247 | 0.049 | 1.020 | 0.262 | 0.228 | 0.904 | 0.261 | 0.237 | 0.891 |
| $\sigma$ | 0.010 | 0.016 | 0.044 | 0.011 | 0.042 | 0.053 | 0.011 | 0.046 | 0.057 |
| $L = 300$ | 0.236 | 0.045 | 0.973 | 0.250 | 0.225 | 0.855 | 0.253 | 0.238 | 0.856 |
| $\sigma$ | 0.010 | 0.015 | 0.043 | 0.010 | 0.042 | 0.047 | 0.011 | 0.042 | 0.054 |
| $L = 320$ | 0.221 | 0.043 | 0.911 | 0.241 | 0.226 | 0.816 | 0.241 | 0.231 | 0.812 |
| $\sigma$ | 0.010 | 0.012 | 0.045 | 0.010 | 0.044 | 0.045 | 0.010 | 0.046 | 0.047 |
| $L = 340$ | 0.211 | 0.042 | 0.871 | 0.231 | 0.219 | 0.778 | 0.231 | 0.233 | 0.767 |
| $\sigma$ | 0.010 | 0.013 | 0.044 | 0.011 | 0.037 | 0.043 | 0.010 | 0.041 | 0.046 |

Such deviations from ideal randomness must be considered in highly precise work. The model which we shall construct, however, is not highly precise; and more sophisticated random number generators designed to remove these imperfections do not give results with our model that are significantly different than those obtained with a standard system generator. Since our purpose is to set up a "reasonably" random standard that anyone can reproduce with readily available computational machinery, we have used a standard congruential generator, the CDC RANF generator for a CDC 6000 series machine.

To transform a standard random number generator into a random protein source, we divide the interval between zero and one into 20 equal parts. When the random number returned by the generator falls into segment one, amino aicd one is placed in a generated sequence, in segment two, amino acid two, etc. With this device we may generate random protein sequences of any length.

If we generate sequences of a given length about 100 times and calculate $R$, $D_1$ and $D_2$ for each sequence, the mean values, $\bar{R}$, $\bar{D}_1$ and $\bar{D}_2$, of the 100 tries become reasonably stable. Table 1 shows the means and standard deviations for these parameters as a function of length. At about length 100 the parameters become stable to about two decimal places for different generator runs and the differences for shorter sequences are not large. The standard deviations for 500 tries are not significantly different than for 100 tries which indicates reasonable convergence at 100 tries. All parameter values are normally distributed over the ensemble of sequences from the random source. If we divide each distribution into increments increasing in size by $0.5\ \sigma$, the observed distribution is not significantly different from normal (chi square $\approx$ one for $v = 7$ for all parameters).

Fig. 1 is a plot of $\bar{R}$ versus length $L$ and Fig. 3 shows $\bar{D}_1$ and $\bar{D}_2$ versus $L$. As the length increases, $\bar{R}$ and $\bar{D}$ decrease.

This is because the general divergence, $D_n$, is the divergence from equiprobability of the $n$-tuples *over and above* that mandated by the lower level divergences, $D_{n-1}$ $D_{n-2} \ldots D_1$. Short sequences are not sufficiently long for all $n$-tuples to be represented equally in the distribution, hence $D_n$ and $R^{(m)}$ decrease with increasing length and approach zero asymptotically for very long sequences.

## Real Sequence Parameters

Classical information theory deals with very long sequences. The length is assumed to approach infinity and by the law of large numbers, the counted frequencies approach probabilities. But protein sequences are very short relative to the size of the alphabet. There are 20 amino acids, hence 400 possible doublets, yet most proteins are less than 500 amino acids long. Clearly even the counted doublet frequencies are not true probabilities; hence it appears, at first thought, mathematically questionable even to attempt the calculation of information theoretic parameters which are functions of true probabilities. However, consider the following.

Let us count the doublet connecting the last and first letter of the sequence. For example, for the sequence, 1–3–6–8–9–5–14–9, we would count the doublet, 9–1. We may now regard the sequence as infinitely long with a repeating unit and the counted $n$-tuple frequencies as probabilities. We may calculate the random standard with or without the connecting doublet and the real sequence parameters in the same way. The consistency criterion is that we compare real and theoretical values from the same model. Numerically it makes no significant difference which model we use after about length 50, and the differences for even shorter sequences are not large.

We have calculated all values counting the connecting $n$-tuples for the sake of mathematical consistency with established information theory. The ergodic model is used throughout. However, if it is argued that proteins are not truly infinitely long sequences, then one could omit the connecting $n$-tuple without significantly affecting the results. It is this insensitivity of the result to the mathematical model used which makes our approach justifiable both biologically and mathematically, provided the consistency criterion is observed.

## Results

The solid curves in Figs. 1 and 3 are theoretical mean values for 100 tries from the random protein generator and the points are calculated values of real protein sequences. Table 2 shows a complete tabulation of all experimental values calculated and the data sources. Dayhoff (1972) and Jukes and Holmquist (1973a) are data compilations and should be consulted for original references.

We have examined 112 protein sequences ranging from length 27—500. Essentially, the first protein of a given length in Table 2 is plotted in all figures. Fragments are not considered, only proteins of well-defined lengths.

Let us define the word-symbol, $L$-random, to mean that the source is random except for the constraint that it emit only sequences of a specified length $L$. This is a precise description of our protein generator.

As $L$ approaches infinity, $R^{(m)}$ and $D_n$ approach zero asymptotically. The criterion, $R^{(m)}=0$, $D_n=0$, defines the state of "ideal" randomness; and in the asymptotic regions this ideal randomness is approached, the rapidity of the approach being a function of the size of the alphabet. However, in the pre-asymptotic regions for shorter $L$ the divergence from zero of the $R$ and $D$ values is a function of $L$. If we find a statistically significant number of experimentally observed sequences outside the standard acceptance interval, $\pm 2\sigma$, of the $L$-random curves, this indicates some additional constraint (beyond short length) acting upon the source which produced these sequences. Sequences which appear outside $\pm 2\sigma$ of the $L$-random curves have rare or unusual information parameters which are quite different from the average sequence produced by the source. Statistically we would expect 5 out of 100 sequences of this type in the output of every $L$-random source.

## The $\bar{R}$ Curves

In Fig. 1 the experimental $R$ values follow the theoretical $L$-random $\bar{R}$ curve in a general way. However, the points tend to lie below the $L$-random curve for $L<175$ and above it for $L>175$. Also 19 values out of our sample of size 112 fall outside $\pm 2\sigma$ of the $L$-random curve which is significantly more than the 6 expected on the basis of chance. Let us select a random sub-sample of size 58 from Table 2 consisting of the first protein of a given length to avoid the charge that our sample is biased due to the presence of homologous proteins. We now expect three proteins out of 58 outside $\pm 2\sigma$. We observe nine, which is still significantly more than the expected value. The ordering within a given length class is purely random and was established before the statistical analysis was even begun. Thus no matter how we select our sample, we observe significantly more than 5% of the $R$ values outside $\pm 2\sigma$. We conclude that there is some additional constraint (beyond short length) acting on the protein source.

Table 2. Experimental protein information parameters

| Protein | $L$ | $R$ | $D_1$ | $D_2$ |
|---|---|---|---|---|
| 1. Secretin, pig | 27 | 0.719 | 1.030 | 2.078 |
| 2. Glucagon, pig | 29 | 0.759 | 0.508 | 2.771 |
| 3. Thyrocalcitonin, pig | 32 | 0.752 | 0.518 | 2.733 |
| 4. ACTH, pig | 39 | 0.705 | 0.467 | 2.579 |
| 5. ACTH, sheep | 39 | 0.679 | 0.579 | 2.355 |
| 6. Posterior pituitary peptide, bovine | 48 | 0.679 | 0.500 | 2.434 |
| 7. Rubredoxin, *Peptostreptococcus elsdenii* | 52 | 0.603 | 0.546 | 2.059 |
| 8. Rubredoxin, *Micrococcus aerogenes* | 53 | 0.658 | 0.401 | 2.444 |
| 9. Ferredoxin, *Micrococcus aerogenes* | 54 | 0.597 | 0.803* | 1.778 |
| 10. Ferredoxin, *Clostridium pasteurianum* | 55 | 0.583 | 0.720* | 1.801 |
| 11. Ferredoxin, *Clostridium butyricum* | 55 | 0.577 | 0.769* | *1.725 |
| 12. Toxin α, Black mamba | 62 | 0.581 | 0.449 | 2.062 |
| 13. Neurotoxin II, North African scorpion | 64 | 0.617* | 0.333 | 2.334 |
| 14. $\alpha_1$-acid glycoprotein, human plasma | 70 | 0.542 | 0.356 | 1.985 |
| 15. Soybean Trypsin inhibitor | 71 | 0.519 | 0.573 | 1.668 |
| 16. Toxin γ, Black mamba | 72 | 0.543 | 0.378 | 1.968 |
| 17. Cytochrome *c551, Pseudomonas fluorescens* | 82 | 0.515 | 0.330 | 1.894 |
| 18. Haptoglobin alpha 1, human F and S | 83 | 0.509 | 0.358 | 1.843 |
| 19. Proinsulin, pig | 84 | 0.517 | 0.450 | 1.784 |
| 20. Cytochrome 85 | 85 | 0.507 | 0.358 | 1.833 |
| 21. Lipoprotein β and γ, sheep | 90 | 0.501 | 0.432 | 1.732 |
| 22. Ferredoxin, *Scenedesmus* sp. | 96 | 0.476 | 0.413 | 1.644 |
| 23. Ferredoxin, alfalfa | 97 | 0.514* | 0.367 | 1.852 |
| 24. Ferredoxin, spinach | 97 | 0.528* | 0.336 | 1.945 |
| 25. Histone IV, bovine and pea | 102 | 0.454 | 0.580* | *1.382 |
| 26. Cytochrome *C558, Euglena gracilis* | 102 | 0.455 | 0.248 | 1.718 |
| 27. Cytochrome *c*, tuna | 103 | 0.448 | 0.283 | 1.652 |
| 28. Ribonuclease T1, *Aspergillus oryzae* | 104 | 0.454 | 0.402 | 1.559 |
| 29. Cytochrome *c*, human | 104 | 0.447 | 0.360 | 1.573 |
| 30. Cytochrome *c*, rhesus monkey | 104 | 0.447 | 0.360 | 1.573 |
| 31. Cytochrome *c*, horse | 104 | 0.436 | 0.435 | 1.449 |
| 32. Cytochrome *c*, donkey | 104 | 0.446 | 0.390 | 1.539 |
| 33. Cytochrome *c*, bovine | 104 | 0.452 | 0.391 | 1.561 |
| 34. Cytochrome *c*, dog | 104 | 0.452 | 0.396 | 1.557 |
| 35. Cytochrome *c*, California grey whale | 104 | 0.448 | 0.387 | 1.549 |
| 36. Cytochrome *c*, Rabbit | 104 | 0.442 | 0.362 | 1.551 |
| 37. Cytochrome *c*, kangaroo | 104 | 0.455 | 0.432 | 1.532 |
| 38. Cytochrome *c*, chicken | 104 | 0.448 | 0.335 | 1.600 |
| 39. Cytochrome *c*, king penguin | 104 | 0.456 | 0.350 | 1.622 |
| 40. Cytochrome *c*, Peking duck | 104 | 0.451 | 0.341 | 1.607 |
| 41. Cytochrome *c*, pigeon | 104 | 0.445 | 0.367 | 1.556 |
| 42. Cytochrome *c*, snapping turtle | 104 | 0.443 | 0.393 | 1.523 |
| 43. Cytochrome *c*, rattlesnake | 104 | 0.476 | 0.356 | 1.702 |
| 44. Cytochrome *c*, bull frog | 104 | 0.454 | 0.341 | 1.623 |
| 45. Cytochrome *c*, dogfish | 104 | 0.461 | 0.319 | 1.674 |
| 46. Cytochrome *c*, lamprey | 104 | 0.459 | 0.325 | 1.658 |
| 47. Cytochrome *c3, Desulfovibrio vulgaris* | 107 | 0.450 | 0.521* | 1.424 |
| 48. Cytochrome *c*, fruit fly | 107 | 0.451 | 0.339 | 1.612 |
| 49. Cytochrome *c*, screw worm fly | 107 | 0.453 | 0.332 | 1.626 |
| 50. Cytochrome *c*, silkworm moth | 107 | 0.439 | 0.301 | 1.596 |
| 51. Cytochrome *c*, tobacco horn worm moth | 107 | 0.443 | 0.300 | 1.616 |
| 52. Cytochrome *c*, *Neurospora crassa* | 107 | 0.435 | 0.353 | 1.528 |

Table 2 (continued)

| Protein | $L$ | $R$ | $D_1$ | $D_2$ |
|---|---|---|---|---|
| 53. Cytochrome *c*, *Ustilago sphaerogena* | 107 | 0.447 | 0.353 | 1.578 |
| 54. Cytochrome *c*, yeast | 108 | 0.442 | 0.269 | 1.643 |
| 55. Thioredoxin, *Escherichia coli* B | 108 | *0.408 | 0.367 | 1.396 |
| 56. Cytochrome *c*, *Candida krusei* | 109 | 0.423 | 0.244 | 1.585 |
| 57. Cytochrome *c*, *Humicola lanuginosa* | 111 | 0.449 | 0.338 | 1.602 |
| 58. Cytochrome *c*, *Crithidia oncopelti* | 111 | 0.452 | 0.318 | 1.634 |
| 59. Cytochrome *c*, mung bean | 111 | 0.443 | 0.256 | 1.657 |
| 60. Cytochrome *c*, castor bean | 111 | 0.436 | 0.284 | 1.602 |
| 61. Cytochrome *c*, sesame | 111 | 0.433 | 0.265 | 1.608 |
| 62. Cytochrome *c*, sunflower | 111 | 0.431 | 0.282 | 1.581 |
| 63. Cytochrome *c*, cotton | 111 | 0.427 | 0.287 | 1.559 |
| 64. Cytochrome *c*, pumpkin | 111 | 0.437 | 0.284 | 1.607 |
| 65. Cytochrome *c*, abutilon | 111 | 0.428 | 0.288 | 1.562 |
| 66. Cytochrome *c*, buckwheat | 111 | 0.427 | 0.297 | 1.551 |
| 67. Cytochrome *c*, cauliflower | 111 | 0.436 | 0.319 | 1.567 |
| 68. Cytochrome *c2*, *Rhodospirillum rubrum* | 112 | 0.427 | 0.433 | 1.412 |
| 69. Cytochrome *c*, wheat | 112 | 0.442 | 0.297 | 1.613 |
| 70. Hemerythrin, sipunculid worm | 113 | 0.441 | 0.187 | 1.717 |
| 71. Ribonuclease, bovine & sheep | 124 | *0.392 | 0.275 | 1.421 |
| 72. Ribonuclease, rat | 127 | 0.412 | 0.200 | 1.583 |
| 73. Azurin, *Pseudomonas fluorescens* | 128 | 0.409 | 0.225 | 1.543 |
| 74. Arginine-lysine-rich histone, calf thymus | 129 | 0.408 | 0.531* | *1.233 |
| 75. Azurin, *Bordetella bronchiseptica* | 129 | 0.407 | 0.309 | 1.450 |
| 76. Lysozyme, chicken | 129 | 0.415 | 0.252 | 1.540 |
| 77. Coat protein, MS2 phage | 129 | 0.426 | 0.293 | 1.550 |
| 78. Histone III, calf thymus | 135 | 0.380 | 0.408 | 1.236 |
| 79. Hemoglobin $\alpha$, human | 141 | 0.395 | 0.471* | 1.236 |
| 80. Hemoglobin $\alpha$, rhesus monkey | 141 | 0.387 | 0.473* | *1.201 |
| 81. Hemoglobin $\alpha$, mouse | 141 | 0.376 | 0.438* | *1.189 |
| 82. Hemoglobin $\alpha$, rabbit | 141 | 0.396 | 0.335 | 1.375 |
| 83. Hemoglobin $\alpha$, horse | 141 | 0.389 | 0.490* | *1.191 |
| 84. Hemoglobin $\alpha$, bovine | 141 | 0.393 | 0.536* | *1.162 |
| 85. Hemoglobin $\alpha$, sheep | 141 | 0.389 | 0.490* | *1.190 |
| 86. Hemoglobin $\alpha$, carp | 142 | 0.391 | 0.346 | 1.345 |
| 87. Hemoglobin $\beta$, bovine | 145 | 0.369 | 0.389 | 1.207 |
| 88. Hemoglobin $\beta$, sheep-fetal | 145 | 0.383 | 0.408 | 1.246 |
| 89. Hemoglobin $\beta$, human | 146 | 0.364 | 0.391 | *1.183 |
| 90. Hemoglobin $\beta$, rhesus monkey | 146 | *0.352 | 0.383 | *1.139 |
| 91. Hemoglobin $\beta$, rabbit | 146 | 0.376 | 0.394 | 1.232 |
| 92. Hemoglobin $\delta$, human | 146 | 0.385 | 0.346 | 1.318 |
| 93. Hemoglobin $\gamma$, human | 146 | 0.398 | 0.275 | 1.444 |
| 94. Nuclease, *Staphylococcus aureus V8* | 149 | 0.390 | 0.329 | 1.356 |
| 95. Myoglobin, whale | 153 | 0.374 | 0.389 | 1.226 |
| 96. Myoglobin, horse | 153 | 0.361 | 0.389 | 1.170 |
| 97. Coat protein, TMV vulgare | 158 | 0.357 | 0.363 | 1.178 |
| 98. Lactoglobulin, bovine AB | 162 | 0.358 | 0.291 | 1.256 |
| 99. Lysozyme, T4 and T2 | 164 | 0.378 | 0.222 | 1.411 |
| 100. Growth hormone, human | 188 | 0.350 | 0.266 | 1.248 |
| 101. Bence Jones $\lambda$, human SH | 213 | 0.304 | 0.307 | 1.006 |
| 102. Bence Jones $\kappa$, human CUM | 221 | 0.299 | 0.257 | 1.036 |
| 103. Trypsinogen, bovine | 229 | 0.320 | 0.351* | 1.032 |
| 104. Elastase, pig | 240 | 0.277 | 0.288 | 0.909 |

Table 2 (continued)

| Protein | $L$ | $R$ | $D_1$ | $D_2$ |
|---|---|---|---|---|
| 105. Chymotrypsinogen A, bovine | 245 | 0.291 | 0.304 | 0.955 |
| 106. Penicillinase, *Staphylococcus aureus* | 257 | 0.305 | 0.451* | 0.869 |
| 107. Carbonic anhydrase, human erythrocyte | 260 | 0.249 | 0.206 | 0.868 |
| 108. Tryptophan synthetase α, *E. coli* K12 | 267 | 0.277 | 0.311 | 0.888 |
| 109. Subtilisin, *Bacillus amyloliquefaciens* | 275 | 0.289 | 0.476* | *0.771 |
| 110. Thermolysin, *Bacillus thermoproteolyticus* | 314 | 0.247 | 0.325* | 0.742 |
| 111. Glyceraldehyde 3PD, pig | 332 | 0.245 | 0.232 | 0.828 |
| 112. Glutamate dehydrogenase, bovine liver | 500 | 0.163 | 0.159 | 0.544 |

Data sources as follows:

Dayhoff (1972), 1–11, 13, 17–25, 27–52, 54–56, 68–73, 75–77, 79–96, 97, 99–106, 108, 109, 111.
Jukes and Holmquist (1973a), 12, 14–16, 53, 74, 78, 98, 107, 110.
Pettigrew (1972), 58.
Pettigrew (1973), 26.
Morgan *et al.* (1972), 57.
Boulter *et al.* (1972), 59–67.
Moon *et al.* (1972), 112.

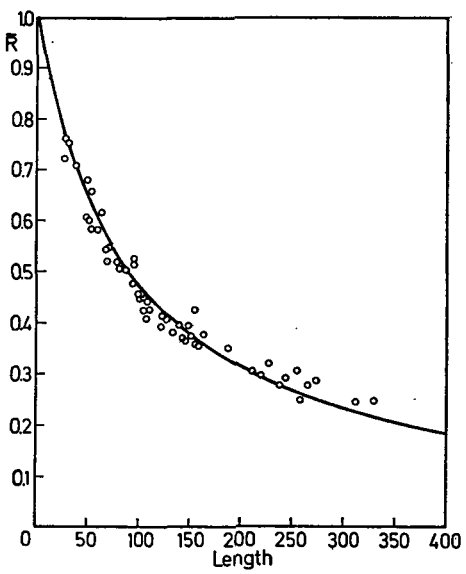



Fig. 1. $\bar{R}$ versus length of sequence for the protein $L$-random source. The solid line is the random theoretical curve and the points are experimental values

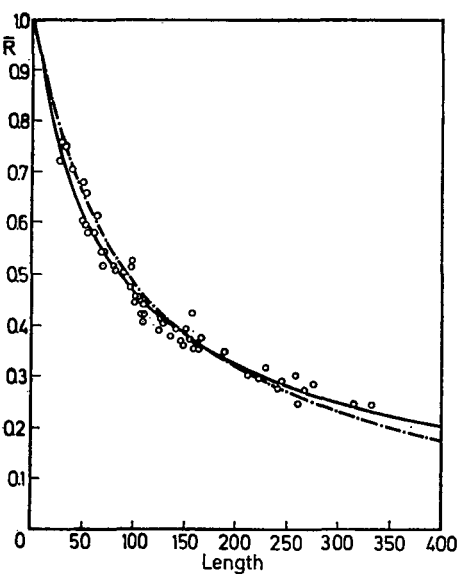



Fig. 2. $\bar{R}$ versus length of protein sequence for the protein $L$-random source (dashed curve) and the DNA-to-Protein $LC$-random code curve (solid curve)

Clearly, present-day proteins are not directly the output of an $L$-random protein source but are produced from DNA sequences via the genetic code. Let us assume a DNA source, random except for the exclusion of the STOP codons and modify our game as follows.

In this game we toss a fair tetrahedral die $3L$ times, replace any STOP codons occurring, and map the resulting DNA sequence to a protein sequence via the genetic code. We repeat this procedure 100 times for a given $L$ and calculate $\bar{R}$, $\bar{D}_1$ and $\bar{D}_2$ for the protein sequence. Since the DNA of this source has a base composition of 50% (C + G), $\bar{R}$, $\bar{D}_1$ and $\bar{D}_2$ and their standard deviations are listed in Table 1 under this heading.

In Fig. 2 the original $L$-random curve from the protein source is now dashed and the solid curve is the $\bar{R}$ curve from the above DNA-to-Protein source. The source is now under the constraint of the genetic code minus the STOP codons as well as the length constraint, but is otherwise random. We will describe this source as $LC$-random and refer to its information parameter functions as "code curves".

The $\bar{R}$ code curve (the solid curve in Fig. 2) lies below the original $L$-random protein $\bar{R}$ curve for $L < 180$ and above it for $L > 180$. This explains the previously observed bias in the experimental $R$ values. At $L = 180$ the curves cross and $R$ is conserved exactly. This is a fundamental symmetry property of the genetic code.

All the experimental points except the 6 whose $R$ values are starred in Table 2 lie with $\pm 2\sigma$ of the $\bar{R}$ code curve. Three out of our sub-sample of size 58 are outside $\pm 2\sigma$. This is no more than we would expect on the basis of chance. In Table 2 values below $2\sigma$ are indicated by a star to the left of the value, and those above $2\sigma$ to the right of it.

We conclude that the redundancy values of proteins as a whole are explained quantitatively by an $LC$-random source, i.e., any non-randomness in the $R$ values of proteins expressed as a divergence from the ideally random value of $R = 0$ is explained quantitatively as due to the short length of the protein plus the constraint of the genetic code.

I regard this result as the expression of a conservation law which states that proteins are the output of a source whose redundancy is random except for the length and code constraints and this redundancy is conserved throughout the evolutionary process. For example, from Table 2 we can predict that 95% of a random sample of proteins of length 340 will have $R$ values in the range of $0.231 \pm 0.022$. Similar predictions can be made for each length value.

## Periodic Proteins

The $R$ conservation principle as stated above has built into it the concept that proteins may be regarded as sequences of well-defined length. This is true for most proteins. However, a class of periodic proteins exists, such as collagen, silk, etc. (*cf.* Yčas, 1972), whose lengths may vary significantly depending on the functional requirements of the particular organism. These proteins perform "mechanical" rather than enzymatic functions. They

consist of a basic repeat unit which was probably originally the output of a random source, but has been repeated an indefinite number of times throughout the length of the protein and modified by point mutation such that the basic repeat unit may no longer be easy to recognize as, for example, in the keratins. This basic repeat unit may be very short, as little as two to five amino acids long.

If a segment of such short length is emitted from a random source, it will have a high $R$ value (*cf.* Fig. 1 or 2). If an indefinite number of such segments are joined together mechanically, the $R$ value will not change as the length increases indefinitely. This is because all the information necessary to calculate any information parameter we might define is contained within the basic repeat unit. The situation is analogous to a bookbinder's error that has bound every page of one kind in a book (Cherry, 1957). Once the first page is read, no further information can be transmitted. In mathematical terms, the $p_{i_\sigma}$ of Eq. (3) are fixed by the basic repeat unit, and these probabilities do not change no matter how many times the unit is repeated.

Therefore the $R$ values of periodic proteins will not follow the $\bar{R}$ code curves which apply only to proteins whose length has been extended by a stochastic mechanism as described previously in our games. The length of periodic proteins has been extended by a mechanical, repetitive mechanism which is fundamentally different. Repetitive length extension is mathematically equivalent to translation parallel to the abscissa in any of our figures. If a randomization process due to mutation is imposed on this translation, it will tend to drag the translation curves down toward the code curves. In the limit this process becomes mathematically equivalent to stochastic length extension. In highly periodic proteins, however, the randomization process is far from complete; and in general the $R$ values of periodic proteins will tend to be high relative to the stochastic curves if sequences of sufficient length are considered. In fact, our methods of anlaysis may be used as a tool for the detection of periodic proteins and characterization of their degree of periodicity.

Table 3 lists four periodic proteins of significant length. All of them have significantly high $R$ values (relative to the $LC$-random source). Listed below each protein is the proposed basic repeat unit (*cf.* Yčas, 1972). Periodic sequences of the specified length were generated with the X-position random. These sequences may be regarded as ideally periodic standard models for the given periodic pattern.

Collagen is more non-random (has higher $R$) than the ideal standard. This confirms quantitatively the observation that the second and third positions of the 3-tuple contain high amounts of proline.

The three keratins all have lower $R$ values than the standard showing that they fall short of perfect periodicity relative to the chosen standard.

Table 3. Periodic proteins

| Protein | $L$ | $\Delta_R$ | $R$ | $D_1$ | $D_2$ |
|---|---|---|---|---|---|
| 1. Collagen (calf skin) | 268 | | 0.487* | 1.359* | *0.745 |
| Gly–X–X | 268 | +0.105 | 0.382* | 0.884* | *0.766 |
| 2. Keratin SCMK–B2A | 171 | | 0.446* | 0.899* | *1.027 |
| Cys–Cys–X–X–X | 171 | −0.010 | 0.456* | 1.061* | *.0911 |
| 3. Keratin SCMK–B2B | 156 | | 0.429* | 0.841* | *1.011 |
| Cys–Cys–X–X–X | 156 | −0.040 | 0.469* | 1.065* | *0.962 |
| 4. Keratin SCMK–B2C | 151 | | 0.417* | 0.876* | *0.928 |
| Cys–Cys–X–X–X | 151 | −0.058 | 0.475* | 1.074* | *0.980 |
| 5. Collagen (rat) fragment | 51 | | 0.637 | 1.305* | *1.446 |
| Gly–X–X | 51 | +0.058 | 0.579 | 1.116* | *1.388 |
| 6. Keratin (wool) fragment | 61 | | 0.561 | 1.041* | *1.382 |
| Cys–Cys–X–X–X | 61 | −0.038 | 0.599 | 1.203* | *1.387 |

Data sources as follows:
Fietzek *et al.* (1973), 1.
Jukes and Holmquist (1973b), 2 and 6.
Jukes and Holmquist (1973a), 3 and 4.
Dayhoff (1972), 5.

These three keratins may now be ordered quantitatively with respect to their degree of periodicity. Let us define

$$\Delta_R = R^{\text{Obs}} - R^{\text{T}} \tag{9}$$

where $R^{\text{T}}$ is the theoretical $R$ value from the ideally periodic standard and $R^{\text{Obs}}$ is the experimentally observed value. $\Delta_R$ orders the keratins linearly as $\Delta_R\text{SCMK} - \text{B2C} < \Delta_R\text{SCMK} - \text{B2B} < \Delta_R\text{SCMK} - \text{B2A}$. A negative $\Delta_R$ value signifies a lack of perfect periodicity in the fixed positions. The $\Delta_R$ value of smallest magnitude then signifies the closest approach of the observed to the theoretical value. We might say that SCMK −B2A has the highest "degree of periodicity" since it comes closest to the ideally periodic standard. A $\Delta_R$ value of zero signifies perfect periodicity relative to the chosen standard. The freezing point depressing glycoproteins are perfectly periodic sequences with the basic repeat unit, alanine-alanine-threonine (De Vries *et al.*, 1971).

For collagen $\Delta_R$ is positive since $R^{\text{Obs}} > R^{\text{T}}$. This signifies an additional non-randomness in the sequence which could be due to an additional periodicity. If we assume that the original repeat unit was Gly–Pro–Pro, then $\Delta_R = -0.359$. Considerable randomization of collagen has occurred relative to this standard.

## The $\bar{D}_1$ and $\bar{D}_2$ Curves

Shannon's redundancy $R$ is divisible into two "components", $D_1$ and $D_2$ as shown in Eq. (7) (*cf.* Gatlin, 1972). $D_1$ measures the divergence from

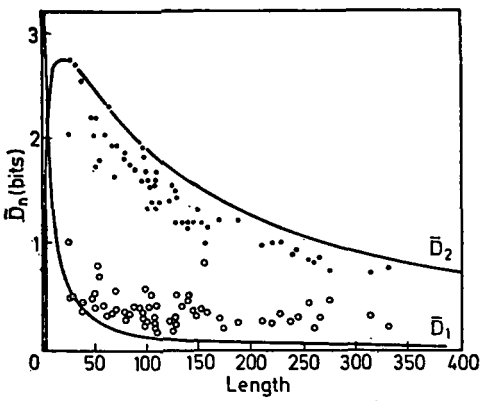



Fig. 3

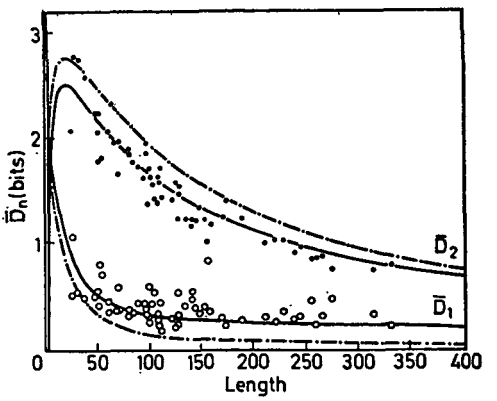



Fig. 4

Fig. 3. $\bar{D}_1$ and $\bar{D}_2$ versus length of sequence for the protein $L$-random source. The solid lines are the theoretical curves. Open circles and solid dots are the experimental values of $D_1$ and $D_2$ respectively

Fig. 4. $\bar{D}_1$ and $\bar{D}_2$ versus length of protein sequence for the $L$-random protein source (dashed curves) and the DNA-to-Protein $LC$-random code curves (solid curves). Open circles and solid dots are the experimental values of $D_1$ and $D_2$ respectively

equiprobability of the single letters in the sequence and $D_2$ measures the divergence from equiprobability of the sequential letter pairs *beyond* that mandated by $D_1$. When we examine the $D_1$ and $D_2$ curves of protein a significant pattern of non-randomness emerges even beyond the length and code constraints.

Fig. 3 shows the $\bar{D}_1$ and $\bar{D}_2$ curves from the original protein $L$-random source before taking the code into consideration. The distributions of experimental values are markedly skewed such that most of the points lie significantly above the $\bar{D}_1$ curve and below the $\bar{D}_2$ curve. In Fig. 4 these $L$-random protein curves are dashed and the solid curves are the code curves.

Again the general bias in the data is explained by the code. The high (relative to $L$-random) $D_1$ values are caused by the asymmetry of the codon assignments; and from Eq. (7) the low (relative to $L$-random) $D_2$ values are explained as the result of $R$ conservation.

However, this time the data are not explained completely and quantitatively by the code and length constraints. Sixteen points lie outside $2\sigma$ of the $\bar{D}_1$ code curve, all of them *above* $2\sigma$, and 11 points lie more than $2\sigma$ *below* the $\bar{D}_2$ curve. These values are starred in Table 2. There are significantly more values outside $\pm 2\sigma$ than we would expect on the basis of chance and their distribution is obviously skewed. From our sub-sample of size 58 there are 9 $D_1$ values above and 4 $D_2$ values below $2\sigma$.

We conclude that a significant tendency exists for the $D_1$ values of proteins to exceed (and the $D_2$ values to fall short of) the limits mandated by short length and the code. This indicates some further constraint upon the DNA source beyond those our calculations have included, some additional non-randomness in the genes.

## Influence of $D_1$ (DNA)

Since randomly generated DNA sequences have a base composition of 50% (C + G), and since all the DNA coding for protein may not have exactly this base composition, we investigate below the influence of change in DNA base composition on protein parameters.

We weighted our tetrahedral die so as to produce a bias toward a base composition of 42% (C + G), the approximate known base composition of all vertebrate DNA. This is accomplished by dividing the interval of the random number generator between zero and one into segments corresponding to the desired base composition. The resulting 42% (C + G) code curves are not significantly different than the *LC*-random or 50% (C + G) code curves. Values are shown in Table 1.

Next we perturbed the Watson-Crick equality of the base composition by about 5% to simulate single strand base frequencies. Again the deviation is well within the statistical error of the average values obtained for double strand base frequencies.

Since the information parameters are not extremely sensitive to changes of this order-of-magnitude in $D_1$ (DNA), it appears likely that the non-randomness in the genes coding for at least some of the unusual proteins contains a $D_2$ component as well. However, the present study does not establish this conclusively.

## Very Short Length Ranges

At $L = 1$ (the monotone), by definition $D_1 = \log a = \log_2 20 = 4.3219$ bits, $D_2 = 0$ and $R = 1$. When we assume a DNA source, random except for the exclusion of the STOP codons and apply the constraint of the genetic code, the maximum value the zero memory entropy of proteins can have is $H_1^{\text{Max}}(P) = 4.1392$ bits. Therefore the theoretical minimum of $D_1$ under these constraints is $D_1^{\text{Min}}(P) = \log a - H_1^{\text{Max}}(P) = 0.1827$ bit. Only 1 value (glutamate dehydrogenase, $L = 500$) in Table 2 is less than this theoretical minimum, and this one value is well within statistical error of the theoretical value.

At $L = 1$, $D_2$ is zero independent of any constraint. $D_2$ rises sharply as shown in both Figs. 3 and 4 reaching a maximum at about $L \cong 20$. The non-coded $D_1$ and $D_2$ curves cross at $L \cong 5$ and the code curves at $L \cong 10$.

## The General *R* and *D* Pattern

The concept of randomness, or its antithesis-organization takes on a multi-dimensional richness in information theory. Classical statistical studies have demonstrated the existence of weak but statistically significant "pair correlations" in proteins (Krzywicki and Slonimski, 1967; Zimmerman *et al.*, 1968; Jones and Dayhoff, 1972). The additional contribution of the present work is to show that, on the average, the divergence from zero of the $D_2$ values of most proteins is explained quantitatively by the length and code constraints. Furthermore, the $D_2$ values of most proteins are actually *less* than that predicted from the length constraint alone. This is because the code imposes a positive $D_1$ constraint and the $D_2$ values are lowered as a result of $R$ conservation. This is considerably different than if proteins were under positive $D_2$ constraint beyond that imposed by their length. *None* of them are.

By contrast, the very long sequences of DNA are under varying positive $D_1$ and $D_2$ constraint, but it is the positive $D_2$ constraint which orders the organisms according to major taxonomic groups, the higher organisms being characterized by the highest optimum values of $D_2$ and $R$ for the *total* DNA (*cf.* Gatlin, 1972). This increase in $D_2$ with evolutionary advancement shows the increase in *fidelity* of the genetic message at the site of primary information storage—the DNA. Protein, on the other hand, does not store the primary hereditary information which is to be transmitted over an information processing channel. Fidelity is not the required value parameter, hence no positive $D_2$ constraint other than the trivial one of length exists. For this reason, $R$ and $D_2$ for proteins are not correlated with evolutionary advancement in the same way as for the total DNA, but reflect rather the structural and functional requirements of the individual proteins.

These conclusions do not agree with those of Reichert and Wong (1971) who claim that $R$ and $D_2$ for proteins are non-random and increase with evolutionary advancement. The present study shows that $R$ and $D_2$ for proteins bear no relationship to evolutionary advancement, but rather are simple functions of the length of the protein. It can be shown that the functions of Reichert and Wong are of the basic form $\sum_i q_i \log p_i$ where the $q_i$ and $p_i$ are elements of two *different* probability distributions. These are completely different functions than my $D_1$ and $D_2$ measures which are of the basic form $\sum_i p_i \log p_i$. It is unfortunate that Reichert and Wong expressed their functions in my notation and claimed analogy to my information measures. A more detailed criticism has been published (Gatlin, 1974).

I thought at first that perhaps $R$ and $D_2$ did increase in the cytochromes *c* of higher organisms because they are on the whole shorter sequences. However, *Euglena gracilis* is a counterexample to this hypothesis because

it is a primitive organism; yet its cytochrome *c* is only of length 102, shorter than any vertebrate. Its $R$ and $D_2$ values surpass those of many vertebrates, human cytochrome *c*, for example. Pages 167–173 of my book (Gatlin, 1972) which were written before the *Euglena* data appeared (Pettigrew, 1973) should be retracted.

## *R* Conservation and the Code

As shown in Fig. 2, at about $L = 180$ the $\bar{R}$ code curve crosses the $L$-random $\bar{R}$ curve of our original protein source. In this region the coded $R$ value is conserved strongly relative to the $L$-random source and to a lesser degree in adjacent regions. The $\bar{D}_1$ code curve is displaced upward throughout its entire length and the $\bar{D}_2$ code curve is displaced downward approaching the $L$-random curve asymptotically with increasing length. For longer lengths, both the $\bar{R}$ and $\bar{D}_1$ code curves are permanently displaced upward while the $\bar{D}_2$ code curve returns to the $L$-random value.

This is all understandable in terms of Eq. (7). An increase in $D_1$ causes an increase in $R$ if $D_2$ is constant. This is the situation for long sequences. However, for short sequences, because the $R$ curves cross, $R$ is conserved and hence $D_2$ values are depressed in short length ranges. Thus the relationships between the information parameters which hold for very long sequences do not necessarily hold for short ones.

All $R$ curves must originate at $R = 1$ for $L = 1$. If the codon assignments are perfectly symmetric, as in the hypothetical code of Table 4, $R$, $D_1$ and $D_2$ are all conserved exactly and follow identically the $L$-random curves. The asymmetry of the codon assignments of the genetic code causes the crossing-over of the $R$ curves. The common origin at $L = 1$ and the crossing-over point tend to hold the curves close together in the length ranges of most common proteins. These purely theoretical considerations suggest the candidacy of $R$ as an evolutionary invariant. They show that once the code is fixed, any large change in $R$ relative to the $L$-random curve would require extremely powerful constraints on the DNA source. Non-random effects of such magnitude do not appear to have played a role in the evolutionary process.

Therefore the general conservative nature of the $R$ parameter is predictable from the code alone. However, the close, statistically significant fit of the experimental data to the $\bar{R}$ code curve is not a priori predictable. These are both important aspects of the general $R$ conservation concept.

Eigen (1971) shows that an error rate of one percent in single digit recognition fidelity restricts reproducible sequences to length about 100. At slightly lower error rates the reproducible sequence length and the $R$ crossing-over point would coincide. Perhaps sequence lengths of this order-of-magnitude were critical in establishing informational stability and identity in early self-instructing systems.

Table 4. A symmetric code

| | | | | |
|---|---|---|---|---|
| Phe | UUC | UCU | CUU | |
| Leu | UUA | UAU | AUU | |
| Ile | UUG | UGU | GUU | |
| Met | CCU | CUC | UCC | |
| Val | CCA | CAC | ACC | |
| Ser | CCG | CGC | GCC | |
| Pro | AAU | AUA | UAA | |
| Thr | AAC | ACA | CAA | |
| Ala | AAG | AGA | GAA | |
| Tyr | GGU | GUG | UGG | |
| His | GGC | GCG | CGG | |
| Gln | GGA | GAG | AGG | |
| Asn | UCA | CAU | AUC | |
| Lys | ACU | CUA | UAC | |
| Asp | UCG | CGU | GUC | |
| Glu | GCU | CUG | UGC | |
| Cys | UAG | AGU | GUA | |
| Trp | GAU | AUG | UGA | |
| Arg | CAG | AGC | GCA | |
| Gly | GAC | ACG | CGA | |
| STOP | UUU | CCC | AAA | GGG |

## The High $D_1$-Low $D_2$ Pattern

The high $D_1$-low $D_2$ values tend to form pairs (occur in the same protein) and in many cases where a high $D_1$ or low $D_2$ value lacks a "mate" the corresponding values are close to the statistical decision point of $\pm 2\sigma$. This high $D_1$-low $D_2$ pattern is an alternative expression of the $R$ conservation principle. If $D_1$ is high due to any cause whatsoever, $D_2$ must be lowered if $R$ is to be conserved [see Eq. (7)].

The high $D_1$-low $D_2$ pattern is exhibited by all the periodic proteins (see Table 3). This signifies the operation of the $R$ conservation principle *within the basic informational unit* of the periodic protein. Hence the most general statement of the $R$ conservation principle is that the redundancy *of the informational unit* of all proteins is conserved at its $LC$-random value. Note that both of the short collagen and keratin fragments have $LC$-random $R$ values (see Table 3).

A simple sign test confirms what is obvious from a visual inspection of Fig. 4, namely that the $D_1$ values tend to lie significantly ($P < 0.05$) above and the $D_2$ values significantly ($P < 0.05$) below the code curves beyond $L = 180$. Therefore the high $D_1$-low $D_2$ pattern becomes more pronounced with increasing length.

Instead of postulating 2 basically different mechanisms of length extension, stochastic and repetitive, as we have done above; one can formulate the highly speculative hypothesis that all present day proteins are the descendents of highly periodic ancestors and differ merely in the extent to which the randomization process has gone to completion. For example, the high $D_1$-low $D_2$ pattern which is so pronounced in the $\alpha$-hemoglobins may be the last remnants of an ancient periodicity which has nearly but not completely disappeared through randomization. The microbial ferredoxins show the same pattern to a lesser extent. There has been previous speculation (Eck and Dayhoff, 1966) that the ferredoxin of *Clostridium pasteurianum* shows evidence of a basic repeat unit of length four.

The tendency for the high $D_1$-low $D_2$ pattern to become more pronounced with increasing length may be interpreted as evidence for a generalized residual periodicity in proteins as a whole. As we explained previously, the high $R$ value of a short segment would tend to remain high relative to the code curves under mechanical length extension. Similar considerations apply to the high $D_1$-low $D_2$ pattern. However, $D_1$ and $D_2$ are more sentitive parameters than $R$ and hence are able to detect the last traces of periodicity for sequences of sufficient length.

Remember that all the high $D_1$-low $D_2$ values starred in Tables 2 and 3 are outside $\pm 2\sigma$ of the code curves and there are significantly more of them than we would expect on the basis of chance. This indicates that the $R$ conservation principle continues to be obeyed even after the $D_1$ and $D_2$ values press beyond the statistical limits of the code.

If some additional constraint on the system is causing $D_1$ to exceed the code limits, there is no a priori reason to expect this new constraint also to obey $R$ conservation. The fact that it does suggests the possibility that $R$ conservation is a "higher" law in the hierarchy of laws governing living systems, and that the code has evolved so as to satisfy $R$ conservation.

On the other hand, if all proteins are indeed descendents of highly periodic ancestors, it is possible that the high $D_1$-low $D_2$ pattern could have been fixed by the code in the short repeat unit and its escape from the code's statistical limits was brought about by nothing more than mechanical length extension. It is possible also that these early periodic "proteins" were not yet coded, at least in the same manner as present-day proteins. Note, for example, that the fibroin of *Bombyx mori* presents many unresolved paradoxes under presently accepted coding concepts (*cf.* Yčas, 1972). At any rate, the code which has evolved is *such that* whenever $D_1$ is raised by the code or any other cause, $D_2$ is reduced accordingly and $R$ is conserved within the basic informational unit.

$R$ conservation is a quantitative detailing of obvious evolutionary principles. Primitive functions form the basis of the hierarchy of controlled functions. Changing a primitive element of the control system is difficult

without restructuring the whole system. Similar concepts have been expressed in biological language (Eck and Dayhoff, 1966). Therefore, it is reasonable to is postulate that conservative principles of many types have governed the evolution of the living system; and Holmquist (1974) has recently discovered electrostatic conservative properties of the code.

Since proteins express the biological meaning of genes, proteins must have acquired their first primitive functional meaning in early self-instructing systems. It is often argued that the genetic code did not start with doublet codons and then switch to triplets, because all of the information acquired up to then would have been destroyed. Similarly, it is obviously advantageous not to destroy all the primitive functional information acquired by proteins. $R$ conservation is a quantitative expression of this principle.

*Acknowledgements.* I wish to thank R. Holmquist, D. W. Hutchinson, C. Antoniac and T. H. Jukes for helpful discussions.

We acknowledge support of NASA grant NGR 05-003-460 to the University of California, Berkeley. I also wish to thank the Department of Applied Science, University of California, Davis, for use of the CDC 3400.

## References

Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M., Brown, R. H.: Proc. R. Soc. Lond. B. **181**, 441 (1972)
Cherry, C.: On human communications. Cambridge, Mass.: Technology Press of Massachusetts, Institute of Technology 1957
Davidson, J. N., Dancis, B. M.: J. Human Evol. **1**, 579 (1972)
Dayhoff, M. O.: Atlas of protein sequence and structure. Silver Spring, Md.: Natl. Biomed. Res. Found 1972
DeVries, A. L., Vander heede, J., Feeney, R. E.: J. Biol. Chem. **246**, 305 (1971)
Eck, R. V., Dayhoff, M. O.: Sci. **152**, 363 (1966)
Eigen, M.: Naturwiss. **58**, 465 (1971)
Fietzek, P. P., Rexrodt, F. W., Hopper, K. E., Kühn, K.: Eur. J. Biochem. **38**, 396 (1973)
Gatlin, L. L.: Information theory and the living system. New York: Columbia University Press 1972
Gatlin, L. L.: J. Mol. Evol. **3**, 233—238 (1974)
Holmquist, R.: J. Mol. Evol. (in press, 1974)
Jones, D. D., Dayhoff, M. O.: Biophys. Soc. Abst. **12**, 162a, SaPM-C11 (1972)
Jukes, T. H., Holmquist, R.: J. Mol. Biol. **64**, 163 (1972)
Jukes, T. H., Holmquist, R.: J. Mol. Evol. **2**, 215 (1973a)
Jukes, T. H., Holmquist, R.: J. Mol. Evol. **2**, 343 (1973b)
Khinchin, A. I.: Mathematical foundations of information theory. New York: Dover 1957
King, J. L., Jukes, T. H.: Sci. **164**, 788 (1969)
Krzywicki, A., Slonimski, P. P.: J. Theoret. Biol. **17**, 136 (1967)
Moon, K., Piskiewicz, D., Smith, E. C.: Proc. Natl. Acad. Sci. (Wash.) **69**, 1380 (1972)
MacLaren, M. D., Marsaglia, G.: J. Assn. Comp. Mach. **12**, 83 (1965)
Morgan, W. T., Hensley, C. P., Jr., Riehm, J. P.: J. Biol. Chem. **247**, 6555 (1972)

Ohta, T., Kimura, M.: Science **174**, 150 (1971)
Pettigrew, G.W.: FEBS Letters **22**, 64 (1972)
Pettigrew, G. W.: Nature **241**, 531 (1973)
Reichert, T. A., Wong, A. K. C.: J. Mol. Evol. **1**, 97 (1971)
Shannon, C. E.: The mathematical theory of communication. Urbana: University of Illinois Press 1949
Yčas, M.: J. Mol. Evol. **2**, 17 (1972)
Zimmerman, J. M., Eliezer, N., Simha, R.: J. Theoret. Biol. **21**, 170 (1968)

Dr. Lila L. Gatlin
Space Sciences Laboratory
University of California
Berkeley, California 94720, USA

# ON THE POSSIBLE ORIGIN AND EVOLUTION OF THE GENETIC CODE

THOMAS H. JUKES

*Space Sciences Laboratory, University of California, Berkeley, Calif. 94720, U.S.A.*

**Abstract.** The genetic code is examined for indications of possible preceding codes that existed during early evolution. Eight of the 20 amino acids are coded by 'quartets' of codons with four-fold degeneracy, and 16 such quartets can exist, so that an earlier code could have provided for 15 or 16 amino acids, rather than 20. If two-fold degeneracy is postulated for the first position of the codon, there could have been 10 amino acids in the code. It is speculated that these may have been phenylalanine, valine, proline, alanine, histidine, glutamine, glutamic acid, aspartic acid, cysteine and glycine. There is a notable deficiency of arginine in proteins, despite the fact that it has six codons. Simultaneously, there is more lysine in proteins than would be expected from its two codons, if the four bases in mRNA are equiprobable and are arranged randomly. It is speculated that arginine is an 'intruder' into the genetic code, and that it may have displaced another amino acid such as ornithine, or may even have displaced lysine from some of its previous codon assignments. As a result, natural selection has favored lysine against the fact that it has only two codons. The introduction of tRNA into protein synthesis may have been a cataclysmic and comparatively sudden event, since duplication of tRNA takes place readily, and point mutations could rapidly differentiate members of the family of duplicates from each. Two tRNAs for different amino acids may have a common ancestor that existed more recently than the separation of the prokaryotes and eukaryotes. This is shown by homology of two *E. coli* tRNAs for glycine and valine, and two yeast tRNAs for arginine and lysine.

## 1. Introduction

The evolution of the genetic code has been a favorite subject of speculation for some years. Two general approaches are often used. The first consists of attempts to find an affinity between amino acids and nucleic acid bases. The idea is to show how a primitive translation system may have started prior to the origin of life. This approach has the advantage of giving rise to laboratory experiments. The second approach, which I shall discuss first, is to work backward from the present code to something simpler. One attraction of this procedure is that it is linked to actual existence, which makes an author feel more secure, like a rock-climber who uses a rope on a cliff, rather than 'climbs free'.

The present code (Table I) has much to commend it. One of its good features is that no one was able to guess what it was, although various people tried to do so. This may mean that no-one will be able to guess how the code started.

One obvious feature of the code that makes one speculate on its possible evolution is its pattern of degeneracy. Eight amino acids: alanine, arginine, proline, glycine, leucine, serine, threonine, valine are coded by two bases plus 'any base in the third position'. An example is the code GCN, for alanine, where N is any base. It is tempting to suggest (Jukes, 1966) that an earlier code, from which the present code is descended, may have been for only 15 amino acids, each coded by a doublet and any third base to act as a 'spacer'.

*Origins of Life* **5** (1974) 331–350. *All Rights Reserved*
*Copyright © 1974 by D. Reidel Publishing Company, Dordrecht-Holland*

TABLE I

The genetic code

| | | | |
|---|---|---|---|
| UUU Phenylalanine | UCU Serine | UAU Tyrosine | UGU Cysteine |
| UUC Phenylalanine | UCC Serine | UAC Tyrosine | UGC Cysteine |
| UUA Leucine | UCA Serine | UAA Chain Termn.. | UGA Chain Termn. |
| UUG Leucine | UCG Serine | UAG Chain Termn. | UGG Tryptophan |
| CUU Leucine | CCU Proline | CAU Histidine | CGU Arginine |
| CUC Leucine | CCC Proline | CAC Histidine | CGC Arginine |
| CUA Leucine | CCA Proline | CAA Glutamine | CGA Arginine |
| CUG Leucine | CCG Proline | CAG Glutamine | CGG Arginine |
| AUU Isoleucine | ACU Threonine | AAU Asparagine | AGU Serine |
| AUC Isoleucine | ACC Threonine | AAC Asparagine | AGC Serine |
| AUA Isoleucine | ACA Threonine | AAA Lysine | AGA Arginine |
| AUG Methionine | ACG Threonine | AAG Lysine | AGG Arginine |
| GUU Valine | GCU Alanine | GAU Aspartic acid | GGU Glycine |
| GUC Valine | GCC Alanine | GAC Aspartic acid | GGC Glycine |
| GUA Valine | GCA Alanine | GAA Glutamic acid | GGA Glycine |
| GUG Valine | GCG Alanine | GAG Glutamic acid | GGG Glycine |

An appealing feature of this suggestion is the finding that the same transfer RNA molecule can be the ancestor of two tRNAs, each for a different amino acid. This is shown by comparing the sequences of *E. coli* glycine and valine tRNAs, and of yeast lysine and arginine tRNAs (Table II). Therefore, there is an existing process for switching the translation of a codon from one amino acid to another. However, there is no way of identifying the amino acid that four codons may have specified in a 'earlier code'. For example, were all four codons that start with AG originally assigned to arginine, or to serine, or perhaps to neither one, since arginine and serine each have four other codons?

## 2. Ambiguous Pairing Between the First Base of Anticodons and the Third Base of Codons

Some anticodons can pair with more than one codon. For example, it was found that the same phenylalanine tRNA would pair with the trinucleotides UUU and UUC; furthermore, no phenylalanine tRNA has been found that will pair with only UUU, or only UUC. When phenylalanine tRNAs were sequenced, the anticodon was found to be GmAA in yeast and wheat (Gm = oxymethyl guanosine) and GAA in *E. coli*. The G or Gm paired with both C and U. This and other findings led Crick (1966) to propose that, while the pairing between the first two bases of the codon and the last two bases of the anticodon followed strictly the regular A·U and G·C rule, there was a certain amount of 'play' or 'wobble' in the pairing between the first base of the anticodon and the third base of the codon. This would be useful biologically, because the utmost fidelity is needed in translating the first two bases of the codon, which are

*This is an offprint of an article which has been published originally in*

# ORIGINS OF LIFE

*An International Journal Devoted to the Scientific Study of the Origin of Life*

Editor: C. Ponnamperuma

One volume is published annually

| | |
|---|---|
| Subscription price per volume of four issues | Dfl. 150,— |
| Plus postage and handling charges | Dfl. 12,50 |
| Total subscription price | Dfl. 162,50 |

Personal subscription price on request

Complete back volumes available

Orders may be placed with your agent or directly with the publisher

## Subscription order form

☐ Please enter my subscription to ORIGINS OF LIFE at the total subscription price of Dfl. 162,50 per volume, as of Volume _____, until cancellation

☐ Check enclosed

☐ Please bill me

Name: _____ Date: _____

Address: _____

_____

Signature: _____

**D. REIDEL PUBLISHING COMPANY**

**DORDRECHT-HOLLAND / BOSTON-U.S.A.**

TABLE II

Comparison of base sequences (unmodified) of *E. coli* tRNAs Valine$_{2B}$ and Glycine$_3$, and Yeast tRNAs Lysine$_H$, and Arginine$_{3A}$

| | | |
|---|---|---|
| Val EC2B | G C G U U C A \| U A \| G C U C \| A G U U – G G U – U A \| G A G C \| A \| C C A C C \| U U G A C A U | |
| Gly EC3 | G C G G G A A \| U A \| G C U C \| A G U U – G G – – U A \| G A G C \| A \| C G A C C \| U U G C C A A | |
| Val EC2B | G G U G G \| G G G – C \| G G U G G \| U U C G A G U \| C C A A U \| U G A A C G C \| A | % Identical bases = 74 |
| Gly EC3 | G G U C G \| G G G U C \| G C G A G \| U U C G A G U \| C U C G U \| U U C C C G C \| U | |
| Lys YH | G C C U U G U \| U G \| G C G C \| A A U C – G G – – U A \| G C G C \| G \| U A U G A \| C U C U U A A | |
| Arg Y3A | G C G C U C G \| U G \| G C G U \| A A U – – G G – – C A \| A C G C \| G \| U C U G A \| C U U C U A A | |
| Lys YH | U C A \| U A A G N U U \| A G G G G \| U U C G A G C \| C C C C U \| A C A G G G C \| U | % Identical bases = 73 |
| Arg Y3A | U C A \| G A A G A U U \| A U G G G \| U U C G A C C \| C C C A U \| C G A G U G C \| G | |

The vertical lines separate helical and non-helical regions.

strictly specific for 8 of the amino acids (see Table I), but the third base of the codon is ambiguous or 'degenerate' in the codes for 18 of the amino acids. Therefore, it might be 'biologically economical' to use the same tRNA for both phenylalanine codons.

By building models, Crick concluded that (i) the alanine anticodon IGC, in which adenine (in adenosine) has been deaminated to hypoxanthine (in inosine), would pair with GCC, GCU and GCA. (ii) either of the two chainterminating codons UAA and UAG would pair with the anticodon UUA. This is thought to be in an *E. coli* serine suppressor tRNA as a result of a mutation from UGA to UUA in the anticodon. (iii) the anticodon CUA, found in another serine suppressor tRNA as a result of a mutation from CGA, would pair only with UAG. Therefore, I would 'wobble' with C, U or A; U would 'wobble' with A or G, but C in the first anticodon position would be specific for G in the third position in the codon.

True to prediction, inosine has been found in the anticodons only of tRNAs for amino acids that have three codons containing U, C and A in the third position. They were alanine, serine, isoleucine, valine and arginine. Most of these tRNAs are from eukaryotic organisms, but one for arginine is from *E. coli,* the tRNAs of which, in bulk, contain very little I.

A new development was reported by Murao *et al.* (1970). They found that uridine-5-oxyacetic acid in the first position of the anticodon oacUAC in *E. coli* valine tRNA would pair with A, G and U, present in valine codons GUA, GUG and GUU, so that the pairing of G in codons was not specific for C in the anticodon. Furthermore, and in contrast, a number of anticodons were found to contain 2-thiouridine in position 1, pairing only with A (Nishimura, 1972), rather than with A and G as in the case of unmodified U (Table III). Evidently, the first base in the anticodon can be modified to produce either greater or less pairing specificity. I conclude that codon-anticodon pairing has changed during evolution to the extent that modification of the bases in anticodons has altered the specificity of pairing with codons in the directions both of more and less ambiguity.

### 3. The Importance of U·G Pairing in the Evolution of the Code

A striking thing about the code is that eleven amino acids are coded by pairs of codons that end either in a pyrimidine or in a purine, but no amino acid is coded only by a pair of codons one of which ends in a pyrimidine and the other in a purine. To give an example, histidine has the codons CAU and CAC; glutamine, CAA and CAG. Surely this pattern must have some evolutionary significance, because the pattern is a functional one, and is related to codon-anticodon pairing. Codons ending with a pyrimidine will pair with anticodons starting with G. Codons ending with a purine will pair with anticodons starting with U. These two general rules hold except for anticodons that start with modified bases. Such modifications are probably of recent evolutionary origin since they differ in different species.

Crick's proposal shows that this pattern in the code depends on U·G pairing between the first base in the anticodon and the third base in the codon ('1–3 pairing')

TABLE III

First base of anticodons in sequenced tRNAs compared with third base of codons paired

| First base in anticodon | Pairing with | Organisms (E = eukaryotes P = prokaryotes) | In tRNAs for |
|---|---|---|---|
| H (I) | U, C, A | E, P | Ile, Val, Ser, (UCN), Ala, Arg (CGN) |
| oa$^5$U | U, A, G | P | Val, Ser (UCN) |
| G | C, U | E, P | Phe, Leu (CUN), Ile, Val, Tyr, Asp, Arg (CGN), Gly |
| Gm | C, U | E | Phe |
| G† ('Q') | C, U | P | Tyr, His, Asn, Asp |
| U | A, G | P | Gly, Term. |
| s$^2$U | A | E, P | Gln, Lys, Glu |
| U† | A, G(?) | E, P | Leu (UUR), Arg (AGR) |
| C | G | E, P | Leu (CUN, UUR), fMet, Gln, Lys, Trp, Gly |
| ac$^4$C | G | P | Met |
| Cm | G | E | Trp |

*Abbreviations*: H, hypoxanthine; I, inosine; oa$^5$U, uridine-5-oxyacetic acid; Gm, 2′-0-methylguanosine; G†, 'Q-base' (an unidentified guanosine derivative); s$^2$U, 2-thiouridine; U†, unidentified uridine derivative; ac$^4$C, 4-acetylcytidine; Cm, 2′-0′methylcytidine; N, unidentified nucleoside; R, purine.

as follows:

| | | |
|---|---|---|
| Anticodons → | GUG | UUG |
| Codons ← | CAC | AAC |
| | UAC | GAC |
| | His | Gln |

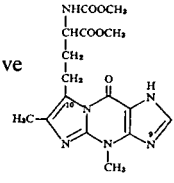
We shall now examine the possibility for U·G pairing between the *third* base in the anticodon and the *first* base in the codon in an earlier genetic code ('3–1 pairing'). It appears that an elaborate system to *prevent* this ambiguity has evolved in tRNA. This is expressed in special modifications of the nucleoside next following the anticodon (site 40) in tRNA (Table IV). There is, in most cases, an isopentenyl-containing side-chain at site 40 when there is 3–1 pairing between A and U, and a threonyl-containing side-chain at this location when there is 3–1 pairing between U and A, and, in prokaryotes, a methyl group when there is 3–1 pairing between G and C. An obvious role for such side-chains is to prevent 3–1 pairing between G and U, and U and G respectively. The prevention of this pairing is necessary to ensure fidelity of translation of codons starting with U or A, *except* in the case of the initiation of polypeptide chains. In *E. coli* the initiator codon is AUG or GUG, either of which is translated as methionine by

TABLE IV

Modifications of the nucleosides at site 40 in tRNAs as correlated with recognition of codons in *E. coli* and other prokaryotes (Prok.) and yeast and other eukaryotes (Euk.)

| Nucleoside adjoining Anticodon | | Nucleoside adjoining Anticodon | | Nucleoside adjoining Anticodon | | Nucleoside adjoining Anticodon | |
|---|---|---|---|---|---|---|---|
| Codon Prok. | Euk. | Codon Prok. | Euk. | Codon Prok. | Euk. | Codon Prok. | Euk. |
| UUY $ms^2i^6A$ | *Y*-base | UCY | $i^6A$ | UAY $msi^6A$ | $i^6A$ | UGY $ms^2i^6A$ | |
| UUA $ms^2i^6A$ | | UCA $ms^2i^6A$ | $i^6A$ | UAA | | UGA – | |
| UUG $ms^2i^6A$ | $m^1G$ | UCG $ms^2i^6A$ | | UAG | | UGG $ms^2i^6A$ | A |
| CUY $m^1G$ | | CCU | | CAY $m^2A$ | | CGY $m^2A$, G | A |
| | | CCC | | | | | |
| CUA | | CCA | | CAA $m^2A$ | | CGA $m^2A$ | A |
| CUG $m^1G$ | | CCG | | CAG $m^2A$ | | CGG | |
| AUY $t^6A$ | $t^6A$ | ACU | | AAY $t^6A$ | | AGY $t^6A$ | |
| | | ACC | | | | | |
| | | | $mt^6A$, | | | | |
| AUA | $t^6A$ | ACA | $t^6A$ | AAA $t^6A$ | $t^6A$ | AGA | $t^6A$ |
| AUG $t^6A$ | | ACG | | AAG $t^6A$ | $t^6A$ | AGG | |
| AUG-f A | $t^6A$ | | | | | | |
| GUY A | A† | GCY | mel | GAY $m^2A$ | $m^1G$ | GGY A | |
| GUA $m^6A$ | A† | GCA A | mel | GAA $m^2A$ | | GGA | |
| GUG $m^6A$ | | GCG | | GAG $m^2A$ | | GGG A | |

Y-base, a guanosine derivative:
$m^1G$, 1-methylguanosine
A†, unidentified adenosine derivative
mel, 1-methylguanosine
$m^2A$, 2-methyladenosine
$m^6A$, 6-methyladenosine
$mt^6A$, N-methylcarbamoyl in $t^6A$



$i^6A$ = 6-($\Delta^2$-isopentenyl) adenosine
$t^6A$ = N-(9-($\beta$-D-ribofuranosyl) purin-6-ylcarbamoyl) -L-threonine[a]
[a] Threonine-containing nucleosides were found in tRNAs accepting arginine and serine
*Y* = U or C
$ms^2i^6A$, 2-methylthio-6-($\Delta^2$-isopentenyl) adenosine

a special tRNA with the anticodon CAU but *without* modification of the adenosine at position 40.* All such modifications take place by enzymatic mechanisms following transcription. These mechanisms can be presumed to be evolutionary innovations. Prior to their appearance, U·G and G·U 3–1 pairing could have taken place in the primitive prokaryotes. Under these conditions, codons starting with G would pair with anticodons ending with either C or U. Therefore the valine codons, GUN, would pair with the eight anticodons represented by NAC and NAU. Similarly, the leucine anticodons represented by NAG would pair with eight codons consisting of CUN and UUN. According to this relationship, isoleucine and phenylalanine would not have been represented in the code. By applying this model throughout codon-anticodon pairing, a code containing ten amino acids is found to be possible (Table V). Eight more amino acids would become members of the code as a result of the suppression of U·G and G·U 1–3 pairing (Table Vb). Two chain terminator codons are introduced

* Note, however, that in eukaryotes this adenosine is modified, and the initiator tRNA pairs with both AUG and GUG.

TABLE V

(a) The ancestral code for ten amino acids, showing anticodon-codon pairing in the first position of codons that includes ambiguity produced by pairing between G and U in addition to G·C and A·U pairing; (b) code 2, produced as a result of elimination of this ambiguity (c) present code resulting from changes in (b). N = U, C, A or G; R = A or G; Y = U or C.

(a) Ancestral Code:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| → Anticodons | NAU | NAG NAA | NAG | NAC | RUU | YUU | RUG RUA | YUG YUA | RUG | YUG | RUC | YUC |
| Codons ← | NUA NUG | NUU | NUC | NUG | YAA YAG | RAA RAG | YAU | RAU | YAC | RAC | YAG | RAG |
| Amino acid | Val | Leu | Leu | Val | Asp | Glu | His | Gln | His | Gln | Asp | Glu |
| → Anticodons | NGU | NGG NGA | NGG | NGC | NCU | NCG NCA | NCG | NCC | | | | |
| Codons ← | NCA NCG | NCU | NCC | NCG | NGA NGG | NGU | NGC | NGG | | | | |
| Amino acid | Ala | Pro | Pro | Ala | Gly | Arg (Orn) | Arg (Orn) | Gly | | | | |

(b) Changes in (a) leading to code 2 as a result of enzymatic modification of tRNA base 40 and consequent suppression of U, G pairing.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| → Anticodons | NAU | RAA | YAA | NAG | NAC | RUU | YUU | RUA | YUA | RUG | YUG | RUC | YUC |
| Codons ← | NUA | YUU | RUU | NUC | NUG | NAA | RAA | YAU | RAU | YAC | RAC | YAG | RAG |
| Amino Acid | Ile | Phe | Leu | Leu | Val | Asn | Lys | Tyr | C.T. | His | Gln | Asp | Glu |
| → Anticodons | NGU | NGA | NGG | NGC | NCU | NCA | NCG | NCC | | | | | |
| Codons → | NCA | NCU | NCC | NCG | NGA | NGU | NGC | NGG | | | | | |
| Amino Acid | Thr | Ser | Pro | Ala | Ser | Cys | Arg (Orn?) | Gly | | | | | |

(c) Changes in (b) leading to present code as a result of reassignment of tRNAs.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| → Anticodons | RAU | NAU UAU | CAU | NCU RCU | YCU | RCA | NCA UCA | CCA |
| Codons ← | YUA | AUA | GUA | YGA | RGA | YGU | AGU | GGU |
| Amino Acid | Ile | Ile | Met | Ser | Arg | Cys | C.T. | Trp |

at this point. The final step is reassignment of several codons, as shown in (c) Table V, so that methionine and tryptophan are added.

To summarize: codon-anticodon pairing at position 1 of the codon may have been ambiguous at an earlier stage in the evolution of the code. The ambiguity disappeared when tRNA evolved to its present specialized state in which the nucleoside at position 40 is often modified, so as to make the pairing specific between position 1 of the codon and position 3 of the anticodon. I therefore postulate that an earlier form of the code could have been as shown in Table V(a). This could reduce the present list of amino acids used in protein synthesis to ten (Jukes, 1973a).

Further support for the possibility of ambiguity in the first position of an early code may be found in the observation that the second base in the codon confers greater specificity on the code than does the first base or, of course, the third base (Woese, 1965). It is therefore logical to attribute more specificity to the second base than to the first base in earlier codes.

The possibility exists that certain amino acids were formerly in the code but have been discarded from protein synthesis as a result of displacement of them from tRNA combination by members of the present 'group of 20'. A mechanism and an example of this was suggested for ornithine. Other amino acids have been found in mixtures obtained by treatment of simpler molecules. These include α-amino-isobutyric acid, α-amino-*n*-butyric acid, β-hydroxyvaline (Miller and Urey, 1959), and β-alanine (Friedman *et al.*, 1972). Whether the ten amino acids in the 'ancestral code' were actually those listed in Table V(a) is open to question in view of the fact that tRNA molecules, as noted above and elsewhere (Jukes and Holmquist, 1972), can 'switch' from one amino acid to another. The main point of the system outlined in Table V(a) is that it would provide for a code in which only ten amino acids would participate.

## 4. Arginine as an Intruder into Evolution

About one-tenth of the codons (6 out of 61) specify arginine. Yet most proteins contain less than 10% arginine, and in most cases the content is far below this. We have examined the arginine content of 83 proteins each containing 50 or more amino acid residues (Table VI). Some of the 83 are representatives or averages of a group of homologous proteins, such as cytochromes *c* or hemoglobins. Others have no homologs that have been sequenced, so these were treated as the sole representatives of their groups. The arginine and lysine contents of the list of Table VI are compared in Table VII. Lysine is great favored over arginine in proteins, in direct contrast to the comparative numbers of codons for these two amino acids. The disparity is less marked in eukaryotes than in prokaryotes. Perhaps the presence of the ornithine cycle makes mammals more tolerant to arginine by providing a mechanism for its breakdown.

The calculation in the last line of Table VII assumes equal amounts of A, U, C and G in messenger RNA for these proteins, formed by transcription of DNA. The DNA of vertebrates contains about 56% A+T and 44% G+C. If these values represent messenger RNA containing an average of 28% A, 28% U, 22% G and 22% C, then, for vertebrates, the expected percentage of lysine codons in all amino-acid codons would be 3.9% rather than 3.3%, and of arginine codons, 9.1% rather than 9.8%.

A protein with less than 5% arginine contains about half the arginine expected from a random distribution of codons (assuming A=C=G=U) and a protein with more than 6.6% lysine contains more than twice the lysine expected from such a random distribution of codons. *Clearly lysine is greatly favored over arginine as a basic amino acid.* The code evidently contains more arginine codons than are needed, and arginine codons in DNA are genetically reduced in numbers by Darwinian selection (King and Jukes, 1969). In contrast, lysine, which has only two codons, actually exceeds arginine

TABLE VI

| Proteins, Eukaryotic | No. of proteins in class | Average | | | % | |
|---|---|---|---|---|---|---|
| | | No. of residues | ARG | LYS | ARG | LYS |
| Trypsinogen, bovine | 1 | 229 | 2 | 15 | 0.9 | 6.6 |
| Lactalbumin | 3 | 123 | 1.3 | 11.7 | 1.1 | 9.5 |
| Ferredoxin – plants | 5 | 96.6 | 1.2 | 4.6 | 1.2 | 4.8 |
| Leghemoglobin – plants | 1 | 140 | 2 | 14 | 1.4 | 10.0 |
| Chymotrypsinogen | 2 | 245 | 4.5 | 12.5 | 1.8 | 5.1 |
| Caseins | 2 | 204 | 5 | 12.5 | 2.5 | 6.1 |
| $\beta$-Lactoglobulin, Bovine AB | 1 | 162 | 3 | 15 | 1.9 | 9.3 |
| Myoglobin | 8 | 152.9 | 2.9 | 19.6 | 1.9 | 12.8 |
| Hemoglobin Alpha Chain | 10 | 141.1 | 3.0 | 11.6 | 2.1 | 8.2 |
| Hemoglobin Beta Chain | 10 | 145.2 | 3.3 | 11.3 | 2.3 | 7.8 |
| *Globin, Chironomus* | 1 | 136 | 3 | 10 | 2.2 | 7.4 |
| Cytochrome *c* | 40 | 107 | 2.4 | 15.6 | 2.2 | 14.6 |
| Protease Inhibitor – Lima Bean | 1 | 84 | 2 | 4 | 2.4 | 4.8 |
| Globin, lamprey | 2 | 146 | 4.5 | 13 | 3.1 | 8.9 |
| Immunoglobuns, light chains: | | | | | | |
| Constant, Kappa, human and mouse | 2 | 106 | 2.5 | 7.5 | 2.4 | 7.1 |
| Constant, Lambda, human and mouse | 2 | 105 | 1 | 7 | 1.0 | 6.7 |
| Variable Kappa, human and mouse | 18 | 107.8 | 4.9 | 3.7 | 4.6 | 3.4 |
| Variable, Lambda, human and mouse | 10 | 109 | 4.0 | 3.2 | 3.6 | 2.9 |
| Immunoglobuns, heavy chain | | | | | | |
| Variable, human | 6 | 109.7 | 5.8 | 4.2 | 5.3 | 3.8 |
| Constant, human | 1 | 336 | 7 | 27 | 2.1 | 8.0 |
| Constant, rabbit | 1 | 326 | 13 | 19 | 3.9 | 5.7 |
| Haptoglobin Alpha Chain | 1 | 84 | 2 | 8 | 2.6 | 10.1 |
| Carbonic anhydrase, human | 1 | 260 | 7 | 18 | 2.7 | 6.9 |
| Hemerythrin, worm | 1 | 113 | 3 | 11 | 2.7 | 9.7 |
| Glyceraldehyde 3-$PO_4$ Dehydrogenase | 2 | 332.5 | 9.5 | 27 | 2.9 | 8.1 |
| Glutamate Dehydrogenase | 1 | 500 | 30 | 32 | 6.0 | 6.4 |
| Thyrotropin $\alpha$ and luteinizing hormone $\alpha$ | 3 | 96 | 3 | 10 | 3.1 | 10.4 |
| Alcohol Dehydrogenase, horse | 1 | 374 | 12 | 30 | 3.2 | 8.0 |
| Muscle triose $PO_4$ isomerase | 1 | 248 | 8 | 20 | 3.2 | 8.1 |
| Chorionic gonadotropin, human $\alpha$ | 1 | 92 | 3 | 6 | 3.3 | 6.5 |
| Adrenodoxin, bovine | 1 | 118 | 4 | 5 | 3.4 | 4.2 |
| Thyrotropin $\beta$ and luteinizing hormone $\beta$ | 2 | 116.5 | 10 | 6.5 | 8.6 | 5.6 |
| Carboxypeptidase, bovine | 1 | 307 | 11 | 15 | 3.6 | 4.9 |
| Prophospholipase $A_2$, pig | 1 | 130 | 5 | 9 | 3.8 | 6.9 |
| Ribonuclease, pancreatic | 3 | 125 | 5 | 9.6 | 4.0 | 7.7 |
| Keratin, High-sulfur protein | 2 | 153.5 | 6.5 | 0 | 4.2 | 0.0 |
| Trypsin inhibitor, pancreatic | 2 | 56 | 2.5 | 3.5 | 4.5 | 6.2 |
| Trypsin inhibitor, soybean | 1 | 71 | 2 | 5 | 2.8 | 7.0 |
| Trypsin inhibitor, ascaris | 1 | 66 | 3 | 7 | 4.5 | 10.6 |
| Cytochrome $B_5$ | 6 | 87.2 | 3.6 | 6.6 | 4.1 | 7.8 |
| Bee venom phospholipase A | 1 | 128 | 6 | 10 | 4.7 | 7.8 |
| Elastase, porcine | 1 | 240 | 11 | 2 | 5.0 | 1.2 |
| Proinsulin | 3 | 83.6 | 4 | 2 | 4.8 | 2.4 |
| Lipotropin $\beta$ and $\gamma$ | 2 | 90.5 | 4.5 | 10 | 5.0 | 11.0 |
| Prolactin | 2 | 194 | 10.5 | 9 | 5.6 | 4.5 |
| Papain | 1 | 212 | 12 | 12 | 5.7 | 4.7 |
| Mouse nerve growth factor | 1 | 120 | 7 | 8 | 5.8 | 6.7 |
| Growth Hormone | 2 | 189.5 | 12 | 10.5 | 6.4 | 5.6 |

*Table VI (Continued)*

| Proteins, Eukaryotic | No. of proteins in class | Average | | | % | |
|---|---|---|---|---|---|---|
| | | No. of residues | ARG | LYS | ARG | LYS |
| Parathyroid Hormone | 1 | 84 | 5 | 9 | 6.0 | 10.7 |
| Neurotoxins, snake venom | 11 | 61.2 | 4.2 | 5.3 | 6.2 | 8.4 |
| Neurotoxin, scorpion | 2 | 63.5 | 2.5 | 5.5 | 3.9 | 8.7 |
| Avidin, chicken | 1 | 128 | 8 | 9 | 6.3 | 7.0 |
| Histone II $B_2$ | 1 | 125 | 8 | 20 | 6.4 | 16.0 |
| Monkey amyloid protein A | 1 | 76 | 4 | 4 | 6.6 | 5.3 |
| Trypsin inhibitor, basic, bovine | 1 | 58 | 6 | 4 | 10.3 | 6.9 |
| $\beta$ chorionic gonadotropin, human | 1 | 139 | 4 | 11 | 7.9 | 2.9 |
| Lysozyme – Vertebrates | 5 | 129 | 11.6 | 6.2 | 9.0 | 4.8 |
| Myelin Membrane Encephalitogenic protein | 1 | 170 | 18.5 | 12.5 | 10.9 | 7.4 |
| Trypsin Inhibitor – Maize | 1 | 65 | 8 | 1 | 12.4 | 1.5 |
| Histone III, Calf thymus | 1 | 135 | 18 | 13 | 13.3 | 9.6 |
| Histone IV | 2 | 102 | 14.5 | 10.5 | 13.7 | 10.8 |
| | Totals | 8909 | 378.5 | 617.8 | Av. 4.25 | 6.93 |
| Proteins, prokaryotic | | | | | | |
| Rubredoxin | 2 | 52.5 | 0 | 3 | 0.0 | 5.7 |
| Cytochrome $C_2$ | 1 | 112 | 0 | 17 | 0.0 | 15.2 |
| Cytochrome $C_{551}$ | 3 | 82 | 0.3 | 8.3 | 0.3 | 10.2 |
| Cytochrome $c_3$ | 1 | 109 | 0.5 | 18.5 | 0.6 | 18.5 |
| 50S Ribosomal protein $A_2$, *E. coli* | 1 | 120 | 1 | 12 | 0.8 | 10.0 |
| Neocarzinostatin *Streptomyces* | 1 | 109 | 1 | 0 | 0.9 | 0.0 |
| Thioredoxin | 1 | 108 | 1 | 10 | 0.9 | 9.3 |
| Ribonuclease $T_1$ | 1 | 104 | 1 | 1 | 1.0 | 1.0 |
| Ferredoxin – clostridial type | 5 | 54.8 | 0.2 | 0.8 | 1.0 | 1.3 |
| Penicillinase *Staph. aureus* | 1 | 257 | 4 | 43 | 1.6 | 16.7 |
| Subtilisin | 2 | 274.5 | 3 | 10 | 1.1 | 3.6 |
| Ferredoxin, *Chromatium* | 1 | 1 | 81 | 2 | 2.5 | 2.5 |
| Azurin | 4 | 125.8 | 1.2 | 12.8 | 1.2 | 9.3 |
| Acyl carrier protein | 1 | 77 | 1 | 4 | 1.3 | 5.2 |
| Coat protein – turnip yellow mosaic virus | 1 | 188 | 3 | 7 | 1.6 | 3.7 |
| Thermolysin, *Bacillus thermoproteolyticus* | 1 | 316 | 10 | 11 | 3.2 | 3.5 |
| Nuclease, staphylococcal | 1 | 149 | 5 | 23 | 3.4 | 15.4 |
| Cytochrome $B_{562}$ | 1 | 110 | 4 | 16 | 3.6 | 14.5 |
| Tryptophan Synthetase A | 1 | 267 | 11 | 13 | 4.1 | 4.9 |
| Aspartate transcarbamylase R Chain | 1 | 152 | 8 | 10 | 5.3 | 6.6 |
| Penicillinase, *B. licheniformis* | 1 | 265 | 15 | 24 | 5.7 | 9.0 |
| Coat Protein – Tobacco Mosaic Virus | 5 | 157.4 | 9.6 | 1.6 | 6.1 | 1.3 |
| Alpha lytic protease, myxobact. | 1 | 198 | 12 | 2 | 6.1 | 1.0 |
| Lysozyme, bacteriophage | 2 | 160.5 | 12.5 | 12.5 | 7.9 | 7.9 |
| | Totals | 3628 | 106.3 | 262.5 | Av. 2.93 | 7.24 |

in most proteins. Lysine ($pK'_3 = 10.5$) confers important basic properties to proteins through its epsilon amino group, but the guanidine group of arginine is far more basic, perhaps excessively so ($pK'_3 = 12.5$). Arginine is formed biochemically from ornithine, ammonia and carbon dioxide via citrulline, or from ornithine and guanidoacetic acid. I suggested (Jukes, 1973b) that ornithine (which is present in peptide linkage in gram-

icidin) preceded arginine in an earlier genetic code and that the arginine codons were originally assigned to ornithine (Table V(a)). Arginine then appeared as a result of the evolution of the urea cycle, and arginine had a greater affinity than ornithine for the then-existent ornithine tRNAs. Such a phenomenon can occur in enzyme chemistry. Aminopterin (4-amino pteroylglutamic acid) has a much greater affinity than the natural substrate, dihydrofolic acid, for the enzyme dihydrofolic acid reductase.

By replacing ornithine in the aminoacylation of tRNAs, arginine displaced ornithine from protein synthesis. This event increased the sophistication of protein molecules by introducing arginine, a new amino acid with unique properties. It replaced an amino

TABLE VII

Distribution of Arginine and Lysine in 83 proteins containing 8909 codons

| | Eukaryotes | Vertebrates only | Prokaryotes |
|---|---|---|---|
| *Arginine* | | | |
| Less than 5 % | 39 | 29 | 19 |
| 5 %–9.9 % | 15 | 15 | 5 |
| 10 % or more | 5 | 4 | 0 |
| *Lysine* | | | |
| Less than 3.3 % | 6 | 4 | 6 |
| 3.3 %–6.5 % | 15 | 10 | 6 |
| 6.6 % or more | 38 | 34 | 12 |
| Arginine: Lysine ratio | 0.613 | 0.613 | 0.405 |
| for random base sequences | 3.0 | 3.0 | 3.0 |

acid (ornithine) which was similar to lysine in the same manner that aspartic acid is similar to glutamic acid, for glutamic acid is homoaspartic acid; lysine is homo-ornithine. The deficiency of ornithine in proteins for needed functions was overcome by increasing the lysine content, accomplished by preferentially selecting lysine codons in evolution. The demand for lysine in the biological synthesis of proteins, incidentally, is so great that lysine deficiency is a major global problem in human nutrition.

It is perhaps surprising that the codons AGA and AGG are not the third and fourth codons for lysine, which has similar codons. Yeast tRNA for arginine, pairing with AGA, is similar to the yeast tRNA for lysine, pairing with AAA (Table II). These two tRNAs differ in their sequences by only 27% (Jukes and Holmquist, 1972). The average difference between pairs of tRNAs for different amino acids involved in protein synthesis is 49.4% ± 7.0 (Holmquist *et al.*, 1973).

If functional constraints are removed from a protein with low arginine and high lysine content, it should evolve in the direction of higher arginine and lower lysine. I suggested in 1969 that most point mutations in the specificity (S) (variable) regions of immunoglobulins are advantageous, and are rapidly incorporated as evolutionary changes, because 'it is immunologically advantageous to have a large available assortment of different antibodies to cope with various antigenic determinants' while 'those

in the constant (C) regions are usually deleterious, thus accounting for the variability of S and the constancy of C sequences,' (Jukes, 1969).

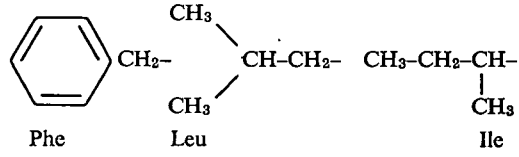
The S and C regions are assumed to have descended from a common ancestor (Hill *et al.*, 1966). The S regions should therefore be higher in arginine and lower in lysine than the C regions as a result of randomly occurring point mutations. Table VI shows that this is indeed the case for both light and heavy human immunoglobulin chains.

## 5. Aspartic and Glutamic Acids

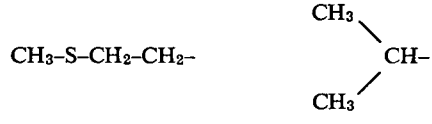
The similarity of the four codons for the dicarboxylic amino acids is frequently held up as an example of the optimization that exists in the code, on the basis that mutations should be non-injurious because of the similarity of these two amino acids. Aspartic and glutamic acids, by 'internally neutralizing' the basic amino acids arginine, lysine and histidine, are important for formation of zwitterions by proteins, and hence for solubility. However, the similarity may be due to chance, because other pairs of amino acids whose codons are identical except for the third base do not have similar chemical properties. These pairs are asparagine and lysine; histidine and glutamine; cysteine and tryptophan; serine and arginine. I therefore conclude that there is no evidence that the similarity of the Asp and Glu codons is necessarily more than a 'fortunate' accident.

## 6. The Hydrophobic Amino Acids; Codons with a Middle U

A conspicuous feature of the code is the similarity in chemical properties of the five amino acids having codons with a middle U. These amino acids have the following hydrophobic side chains:




Phe Leu Ile




Met Val

The relationship between codons and properties is often considered to indicate optimization in that the codons for all these amino acids are interconnected by single base changes in the first position. The effect of this is alleged to be that mutations involving the first or third positions of these sixteen codons might tend to be neutral or near-neutral rather than deleterious. Here we have a hint of a process in evolution favoring neutral, non-adaptive changes.

Methionine has a single codon and a special function in protein synthesis as the amino acid that initiates the synthesis of polypeptides, following which it is removed enzymatically when or before the protein is completed. Methionine has other special functions in the interior of polypeptide sequences. We suggest that the elaborate mechanism by which methionine initiates polypeptide synthesis is an evolutionary development that appeared after the code reached its present from, and that valine preceded methionine as the 'initiator' amino acid.

*Why do some amino acids have six codons*? It was pointed out that so many codons seem unnecessary in the case of arginine. Obviously, however, every possible sequence of three bases in a strand of messenger RNA (and hence in the corresponding DNA sequence) must be translated as an amino acid or serve as a termination signal. There should not be a large number of termination codons in the code, or it will be too difficult to maintain the existence of long polypeptide chains against the frequent occurrence of mutations that produce premature termination, with lethal results. The code therefore cannot contain a high proportion of untranslated codons.

Each codon must contain a minimum of 3 nucleotides to provide for distinguishing between 20 different amino acids.

Most mutations in the third codon position are 'silent' or neutral because they do not change the meaning of the mutated codon. It is also of interest that the six codons for serine are such that each of them can be changed to a codon for the similar amino acid threonine by a single-base change. However, it is not clear why there should be six codons for serine and only two for the much-used glutamic acid. Nor is it clear why there are six codons for leucine and only three for isoleucine. Comparisons of the sequences of tRNAs do not resolve these problems, because the sequences of two tRNAs for *different* amino acids usually differ by about 50% and in most cases, the sequences of two tRNAs for the *same* amino acid in two different organisms also differ by about 50% (Holmquist *et al.*, 1973).

We must answer the question: If at one time the number of amino acids used in protein synthesis was 16 or less, why was the code not 'frozen' as a doublet code? Eigen (1971) answered this on the basis of stability constraints, using data compiled from equilibrium and rate studies with oligonucleotides. He comments:

> The most interesting effect is the preference for the triplet, however, not just for the logically obvious reasons, i.e. the prerequisite for the coding of more than 20 symbols, but rather due to mechanistic coincidences. Codons with less than three digits would be very unstable (at least for A and U). Codons with more than three digits, especially for G and C, become too 'sticky'. The life time of a codon-anticodon pair should not exceed milliseconds so that enzymes with corresponding turnover numbers can adapt optimally. The same type of optimization between stability and rate is always found for enzyme-substrate interactions. Any gain in stability means a lowering of complex dissociation rates; these have to match the turnover numbers in order not to become the rate limiting steps for the turnover.

Gatlin (1972) has found the three-digit codon to be informationally optimal for translation on a theoretical basis. Her calculations show that the entropy $H_1$ for the the protein 'language', $H_1(P)$, reaches its maximum for a 64-variable source when the codons are triplet sequences.

## 7. Evolution of tRNA

The sequenced transfer RNAs are a group that includes tRNAs for 16 of the 20 amino acids. Their sequences are sufficiently similar that we can conclude that they have descended from a single ancestral molecule by gene duplication followed by mutations. This conclusion is reinforced by the fact that pairs of tRNAs for the same amino acid in the same organism may differ by as little as 3% or as much as 36% (Holmquist *et al.*, 1973), showing that the process of duplication and evolutionary separation of tRNAs is a continuous and ongoing process. The inference that the tRNAs have a single ancestor leads to the additional conclusion that the original ancestral tRNA could not have accepted more than one amino acid, or contained more than one anticodon. The evolution of tRNA therefore leads us back to the beginnings of the present genetic code, which depends on recognition of adaptor molecules by the codons present in the genetic message. We cannot switch from a 3-letter code to a one containing only two letters per codon without destroying all the accumulated hereditary information, and, in effect, originating life over again.

## 8. Slippage in DNA Replication

Kornberg *et al.* (1964) discovered that enzymatic synthesis of DNA will take place in the absence of a template to produce repetitive sequences of poly $d(AT)_n$, in the presence of dATP and dTTP as substrates. The process was greatly accelerated by adding oligonucleotides with sequences of 6 to 14 alternating deoxyadenylate and deoxythymidylate residues, such as $d(AT)_6$. The authors postulated that successive stages of replication and slippage resulted in the continuous reiteration of the template and the synthesis of a large dAT co-polymer. This phenomenon was utilized by Khorana *et al.* (1966) in experiments to synthesize repetitive sequences in RNA for solving the problem of the genetic code.

DNA 'families' of identical sequences are recognized as being of comparatively recent evolutionary origin (Britten and Kohne, 1965). Some of these families separate from the main bulk of DNA upon centrifugation and are therefore termed 'satellite' DNA. Satellite DNA of guinea pigs contains a repeating sequence of about six nucleotides (Southern, 1970). Such a molecule would be a possible starting point for coding the synthesis of 'new' polypeptides consisting of repetitions of short oligopeptides that if 'biologically useful', could form functional proteins which could eventually become modified by point mutations. Yčas (1972) has reviewed the existing information on proteins that contain periodic repetitions. Some are remarkably simple, such as a lepidopteran silk fibroin consisting mainly of $(Ala\text{-}Gly)_n$. Ycas points out that collagen has possibly evolved from $(Gly\text{-}Pro\text{-}Pro)_n$, changed by a number of replacements of proline residues by other amino acids.

These repetitive proteins are structural rather than enzymatic in function. They are 'new' in the evolutionary sense, and therefore do not cast light on the composition of earlier genetic codes. We do not know whether enzymes were formed by evolutionary

changes in long, repetitive peptide chains that were originally without catalytic activity. Another possibility, which seems more feasible, is that enzymes started as short peptides with weak catalytic activity, and that these short peptides evolved by lengthening to proteins with greater activity, and with structures that enhanced their stability and specificity.

## 9. Evolution in Transfer RNAs

The genetic code is considered to be universal in terrestrial organisms, and to be 'frozen'. This may be true of the assignments of the codons, but the tRNAs, which function in translation of the code, are evidently still in process of evolution as judged by the presence of duplicate forms of tRNA molecules differing only slightly from each other, such as the two serine tRNAs of yeast (Zachau *et al.*, 1966).

The system that synthesizes polypeptides in cells is extremely complex. It involves ribosomes, which contain large numbers of different proteins as well as RNA. Many enzymes and coenzymes participate in the synthesis of proteins upon ribosomes. The entire process is so elaborate that it appears unlikely to resemble in any respect the primitive systems that produced polypeptides during the period when life was first emerging. A discussion of the origin of ribosomes would be pertinent in an essay on the origin of the code, but, in view of the paucity of information on this subject, I shall not undertake it here.

There are two tRNAs in *Staphylococcus epidermidis* that insert glycine into peptidoglycan molecules, which are units of the bacterial cell wall. These tRNAs differ from the tRNAs that participate in the synthesis of polypeptides upon ribosomes, but the essential differences are only in the absence of *T* and Ψ from loop IV, and U instead of purine at position 40. The remainder of the molecule, including the regular 'cloverleaf' structure, is quite homologous with the other tRNAs (Holmquist *et al.*, 1973). The synthesis of peptidoglycans, however, may well be an evolutionary development that took place subsequently to the appearance of the tRNA-ribosomal system of polypeptide synthesis.

This evolutionary versatility shown by the tRNAs suggests that they could have undergone an epochal event of rapidly-occurring duplication followed by mutations that brought about numerous changes. These changes could have led to the appearance of a large number of different anticodons and different recognition sites for various amino acids.

If this development took place at the same time as the appearance of the ribosomal RNAs, the fundamental units of a polypeptide-synthesizing system, ribosomes and tRNAs, could have made their appearance together.

Holmquist *et al.* (1973) found that pairwise comparisons, of 43 different tRNAs showed that their primary structures had diverged so far that it is impossible to construct a coherent phylogeny for most of the 43. The average divergence, $49.4\% \pm 6.9$, for pairs of tRNAs for different amino acids involved in protein synthesis represents an equilibrium between the restraints of natural selection and the flow of point mutations (Holmquist *et al.*, 1973).

The initial appearance of something resembling the present tRNA on the biological scene must have been a portentous event. It could well have taken place suddenly, because, as Eigen (1971) has shown, a 'clover-leaf' structure emerges from a game played with a random sequence of 80 digits composed of A, U, G and C. Each player throws a tetrahedral die, each face of which represents one of the four letters, and tries to approach a double-stranded structure with as many as possible A·U or G·C pairs. This 'game' in evolution would be played by a system that 'tried to resist' hydrolysis by forming base pairs. Once formed, the tRNA molecule could undergo rapid proliferation by gene duplication, and point mutations would soon produce a series of different anticodons, for the anticodon loop would be exposed in single-stranded form by the very formation of the clover-leaf structure. Such a 'family' of tRNAs could furnish enough anticodons to pair with all the codons, especially if codon-anticodon pairing were only partially specific, as is the case in the third position of codons. If pairing were relaxed in the first as well as in the third position, a pattern could exist as shown in Table V(a). The pattern of amino acid binding by tRNA might develop more slowly than the changes in anticodons; at first, only one or two amino acids might be bound by the primordial tRNAs. Advantages would soon accrue to the system that built up the number of amino acids that were bound by tRNAs and thus were brought into protein synthesis.

The argument that optimization of the code is shown by the nature of amino acid

TABLE VIII

Amino acid interchanges that can occur as the result of single-base changes in the coding triplets

| Amino acid | Possible interchanges[a] |
|---|---|
| Ala | Asp Glu Gly Pro Ser Thr Val |
| Arg | Cys Gln Gly His Ile Leu Lys Met Pro Ser Thr Trp |
| Asn | Asp His Ile Lys Ser Thr Tyr |
| Asp | Glu Gly His Tyr Val |
| Cys | Gly Phe Ser Tyr Trp |
| Gln | Glu His Leu Lys Pro |
| Glu | Gly Lys Val |
| Gly | Ser Trp Val |
| His | Leu Pro Tyr |
| Ile | Leu Lys Met Phe Ser Thr Val |
| Leu | Met Phe Pro Ser Trp Val |
| Lys | Thr Met |
| Met | Thr Val |
| Phe | Ser Tyr Val |
| Pro | Ser Thr |
| Ser | Thr Trp Tyr |

[a] The underlined examples have been reported to occur in mutations.

replacements resulting from single-base changes is unconvincing. The 75 possible amino acid interchanges resulting from single-base changes are in Table VIII. It is obvious that many of them represent interchanges between amino acids of widely differing properties. We conclude that the only optimization shown by the code is the fact that many of the changes in the third base of codons do not produce changes in amino acids. This feature may be an incidental result of the spatial nature of codon-anticodon pairing rather than an 'evolutionary optimization'.

## 10. Origin of a Translation System

The proposal that the genetic code originated by some kind of loose affinity between amino acids and the bases of nucleic acids is an obvious one. Indeed, it was the basis of the first suggestions for a genetic coding system made by Dounce (1952) and Gamow (1954). These suggestions were discarded when they were put to flight by Crick's 'adaptor hypothesis' in 1955. However, although the above proposal is obvious, it runs into very serious practical difficulties. Some of these were discussed by Razska and Mandel (1972). It is fairly easy to show that some amino acid, such as phenylalanine, has a slight preferential affinity for binding weakly to some base, such as uracil, under specialized conditions, but nothing has emerged from these experiments that indicates a special attraction of each individual codon or (just as logically) each anticodon, for the appropriate member of the list of 20 amino acids. This defect is usually countered by saying that even a slight preference in binding, spread over hundreds of millions of years of chemical trial and error, would result in the emergence of a coherent code. Serious objections can be made to this rationalization, because the essence of a successful coding system is *fidelity*, and the errors arising from a pairing mechanism containing a high degree of indefiniteness or ambiguity would be so great as to prevent the emergence of the process of heredity. This objection was raised and discussed by Eigen (1971).

The protagonists of the 'weak binding' theory for the origin of the code include Woese (1967) whose viewpoint was summarized by him in 1967 as follows:

> ...amino acid-nucleic acid 'recognition' interactions (or their equivalent), ... being very weak interactions... could not have given rise directly to the genetic code as it now exists. Instead they must have played the role of constraints operating on the evolutionary process, in this way gradually shaping the form of the code. Since the interactions are weak and therefore cannot manifest an all-or-none sort of specificity with regard to amino acid-codon pairings, it is reasonable to expect that they cannot align amino acids along a nucleic acid template directly, and so this 'recognition' role has been filled by the evolution of an 'intermediary' system, the tRNA's and activating enzymes, that recognize with very high accuracy both an amino acid and its codons.

An objection to this line of reasoning is that it tends to invoke the existence of interactions that are so weak that they cannot be detected except under specialized and artifical experimental conditions. I feel that such an explanation may rely on a desire to postulate non-existent phenomena because we feel intuitively that they ought to have existed since we cannot think of any other way that the genetic coding of proteins might have started. There are infinitely large numbers of ways in which long sequences

of variables can be arranged. The task of natural selection working alone to find useful members of an infinitely large set would be impossible. Therefore, there is a tendency to hope that there must have been an orderliness in aligning amino acid with nucleic acids in some primitive system.

A strong objection to the 'stereochemical fit' concept is that the two sets of codons for serine, UCN and AGY, are so dissimilar. Various authors have proposed that there is a relationship between the second base of codons and the chemical properties of the amino acid, and that this relationship has governed the evolution of the code. Such a relationship, in my opinion, is perceptible only for the codons with U in the middle position. The other three 'groups' (Table I), do not share common properties. For example, although serine and threonine are similar, they do not resemble proline; neither are cysteine, tryptophan and arginine similar to each other. Moreover, serine codons are found in two different classes. The 'clustering' of the hydrophobic amino acids could have come about by the phenomenon postulated in Table V, in which all 16 NUN codons are shown as having evolved through three stages of increasingly

TABLE IX

Relative abundance of certain elements in the Earth's crust and in sea-water (After Mason, 1952; Sverdrup *et al.*, 1942; Bowen, 1966)

| Element | Presence in Earth's crust ppm | Sea-water pp $10^9$ |
|---|---|---|
| Fe | 50000 | 2 to 20 |
| Mn | 1000 | 1 to 10 |
| Cr | 200 | 0.05 |
| Zn | 132 | 5 to 14 |
| Ni | 80 | 0.1 to 5.4 |
| Cu | 70 | 1 to 9 |
| Co | 23 | 0.1 to 0.27 |
| Mo | 15 | 0.3 to 10 |
| Se | 0.09 | 0.09 to 4 |

precise codon-anticodon pairing. The fact that five amino acids with hydrophobic side-chains all have codons with a middle U may have been shaped by evolution, or it may be a coincidence.

Mention should be made of the romantic suggestion by Crick and Orgel (1973) that the genetic code could be of pan-spermatic, extra-terrestrial origin; coming from an extra-galactic source via space-ship. In their proposal, Crick and Orgel go to the length of inferring that the beings who sent us the code dwell in an environment that is rich in molybdenum. Their argument is that terrestrial organisms, in contrast to the Earth's crust, are higher in molybdenum than in nickel. However, the proportions of minerals in animals are thought to resemble the composition of these found in sea-water rather than that of the geosphere. The six trace elements that are components

of identified enzyme systems are iron, zinc, copper, manganese, molybdenum and selenium. A seventh, cobalt, is not utilized as an inorganic element; animals (but not green plants) use it as a component of vitamin $B_{12}$, which they obtain from micro-organisms. It is interesting to compare the six in terms of their abundance in the Earth's crust and in sea-water (Table VI). They show a tremendous range in their abundance in the earth's crust, extending between six and seven orders of magnitude. However, their concentrations in sea-water, in which life is thought to have originated, are remarkably uniform. The average values for each of the six 'enzymatic' trace elements extend from 1 to 11 parts per $10^9$. Note that chromium occurs below this range, and so does cobalt. Molybdenum occurs in sea-water at higher concentrations than nickel or chromium. It is unnecessary to postulate that molybdenum stars might have served as a jumping-off point for pan-spermatic organisms that 'infected' the Earth. In any case, the mineral content of protoplasm evidently is greatly influenced by natural selection, judging from anomalies such as the iodine content of the thyroid gland, the vanadium content of tunicates, etc.

## 11. Summary

(1) Suggestions are made for possible pathways of evolution of the genetic code, assuming that the present code was preceded by codes that contained fewer amino acids. It is recognized, however, that it is possible that earlier codes contained amino acids that are not currently used in protein synthesis.

(2) As an example of the latter phenomenon, it is suggested that ornithine preceded arginine in the code. Support for such a suggestion is found in the observation that the average arginine content of proteins is only about half the value expected from the fact that arginine has six codons. It is proposed that the introduction of arginine led to a diminished use of the codons CGN and AGR, and an evolutionary selection of more lysine resulted than would be expected from the fact that it has only two codons.

(3) Pairing between G and U (or U and G), in the first position of codons and the third position of anticodons could reduce the number of amino acids involved in protein synthesis to ten. It is suggested that this formed the basis of an earlier code.

(4) The introduction of tRNA into protein synthesis may have been a cataclysmic and comparatively sudden event, since duplication of tRNA takes place readily, and point mutations could rapidly differentiate members of the family of duplicates from each other.

## Acknowledgements

This work was supported by NASA Grant NGR 05-003-460 to the University of California, Berkeley.

I wish to dedicate this essay to the memory of Prof. J. B. S. Haldane, who asked me to write a manuscript for him during the meeting on 'The Origins of Prebiological Systems and of their Molecular Matrices' at Wakulla Springs, Florida, November 1963.

## References


Bowen, H. J. M.: 1966, in *Trace Elements in Biochemistry*, Academic Press, London and N.Y.

Britten, R. J. and Kohne, D. E.: 1965–1966, *Carnegie Institute*, Washington, D.C. 20015, Year Book, pp. 78–106.

Crick, F. H. C.: 1966, *J. Mol. Biol.* **19**, 548.

Crick, F. H. C. and Orgel, L.: 1973, *Icarus* **19**, 341.

Dounce, A. L.: 1952, *Enzymologia* **15**, 251.

Eigen, M.: 1971, *Naturwissenschaften* **58**, 465.

Friedman, N., Haverland, W. J., and Miller, S. L.: 1972, in R. Buvet and C. Ponnamperuma (eds.), *Molecular Evolution* **1**, 123, North-Holland, American Elsevier Publishing Co., Inc., New York.

Gamow, G.: 1954, *Nature* **173**, 318.

Gatlin, L. L.: 1972, *Information Theory and the Living System*, Columbia University Press, New York, pp. 1–210.

Hill, R. L., Delaney, R., Fellows, R. E., Jr., and Lebovitz, H. E.: 1966, *Proc. Nat. Acad. Sci. U.S.A.* **56**, 1762.

Holmquist, T., Jukes, T. H., and Pangburn, S.: 1973, *J. Mol. Biol.* **78**, 91.

Jukes, T. H.: 1966, *Molecules and Evolution*, Columbia University Press, New York and London, pp. 1–285.

Jukes, T. H.: 1969, *Biochemical Genetics* **3**, 109.

Jukes, T. H.: 1973a, *Nature* **246**, 22.

Jukes, T. H.: 1973b, *Biochem. Biophys. Res. Commun.* **53**, 709.

Jukes, T. H. and Holmquist, R.: 1972, *Biochem. Biophys. Res. Commun.* **49**, 212.

King, J. L. and Jukes, T. H.: 1969, *Science* **164**, 788.

Khorana, H. G., Buchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kossel, H., Morgan, R., Narang, S. A., Ohtsuka, E., and Wells, R. D.: 1966, *Cold Spring Harb. Symp. Quant. Biol.* **31**, 39.

Kornberg, A., Bertsch, L. L., Jackson, J. F., and Khorana, H. G.: 1964, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 315.

Mason, B.: 1952, *Principles of Geochemistry*, John Wiley and Sons.

Miller, S. L. and Urey, H. C.: 1959, *Science* **130**, 245.

Murao, K., Saneyoshi, F., Harada, F., and Nishimura, S.: 1970, *Biochem. Biophys. Res. Commun.* **38**, 657.

Nishimura, S.: 1972, *Progr. Nucl. Acid. Res. Mol. Biol.* **12**, 49.

Raszka, M. and Mandel, M.: 1972, *J. Mol. Evol.* **2**, 38.

Southern, E. M.: 1970, *Nature* **227**, 794.

Stewart, T. S., Roberts, R. J., and Strominger, J. L.: 1971, *Nature* **230**, 36.

Sverdrup, L., Johnson, W., and Fleming, J.: 1942, *The Oceans* Prentice Hall.

Woese, C. R.: 1965, *Proc. Nat. Acad. Sci. U.S.A.* **54**, 1546.

Woese, C. R.: 1967, *Progr. Nucl. Acid Res. Mol. Biol.* **7**, 107.

Yeas, M.: 1972, *J. Mol. Evol.* **2**, 17.

Zachau, H. G., Dütting, D., and Feldmann, H.: 1966, *Z. Physiol. Chem.* **347**, 212.

DSDQ

BIOPOLYMERS VOL. 13, 345–357 (1974)

# Circular Dichroism of Films of Polynucleotides

WENDEL C. BRUNNER* and MARCOS F. MAESTRE, *Space Sciences Laboratory, University of California, Berkeley, California 94720*

## Synopsis

The CD spectra of films of the lithium salt of *E. coli* and calf thymus DNA, and alternating d-AT:AT were measured as a function of relative humidity. Films of the ammonium acetate salt of DNA were also measured. The ammonium films yield the previously reported A-form CD spectra. A possible explanation for the small magnitude of the 260-nm band of the A-form film spectra compared to double-stranded RNA spectra is that the film DNA is in a different conformation than RNA within the A family of conformations.

At relative humidities of 92% or lower, a negative nonconservative CD spectrum with negative minima near 270 and 210 nm is observed with the lithium films. The magnitude of the minima varies from film to film. In films of DNA the magnitude ranges from a delta epsilon of −5 to −35; d-AT:AT films show magnitudes to −300. CD spectra of this type are designated Ψ spectra. Similar spectra have been reported from reconstituted complexes of DNA and polylysine or f-1 histone. If the origins of the film and protein–DNA complex spectra are similar, the complex spectra are not the result of specific secondary structural changes induced in the DNA by the protein fraction. Theoretical analysis suggests that Ψ spectra are not the result of changes in the secondary or tertiary structure of DNA. Instead, the previously proposed explanation based on liquid crystals is favored. The DNA could form asymmetric structures with long-range periodicity. It is likely that the observed CD spectra of f-1 complexes are artifacts of DNA aggregation. The possibility that some other previously published spectra of protein–DNA complexes also reflect artifacts is suggested.

## INTRODUCTION

One approach to studying nucleic acid-containing biological structures is to determine the conformation of the nucleic acid polymer in the structure. Circular dichroism spectroscopy is a potentially useful tool for studying nucleic acid conformation.[1–3] The interpretation of CD data is difficult however, because at present there is no adequate theory that can relate the composition and conformation of a nucleic acid polymer to its measured CD spectrum. In the absence of such a theory, it is necessary to study the CD of polynucleotides under a variety of physical conditions. The CD spectra of biological structures can be compared to other polynucleotide CD spectra, and in this manner the biological data can sometimes be interpreted.

* Present address: Department of Chemistry, University of Oregon, Eugene, Oreg. 97403.



© 1974 by John Wiley & Sons, Inc.
REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

The present paper reports the CD spectra of films of the lithium salt of *E. coli* DNA, calf thymus DNA, and double-stranded alternating deoxyadenosine and thymidine (d-AT:AT), which have been equilibrated against various relative humidities. The techniques follow those of Tunis-Scheider and Maestre.[4,5] Also reported are the spectra of films of the ammonium salt of *E. coli* DNA. These data should be useful in interpreting many of the results of CD studies of nucleohistone complexes.

## MATERIALS AND METHODS

### DNA Preparation

*E. coli* and calf thymus DNA were obtained from Worthington Biochemical Corporation, and d-AT:AT was obtained as a solution in 0.1 *M* Tris buffer from Biopolymers. *E. coli* and calf thymus DNA were dissolved in about 1.5 ml of $10^{-3}$ or $5 \times 10^{-4}$ *M* LiCl or $10^{-3}$ *M* $NH_4CH_3COO$ to form a very concentrated, viscous solution. The resulting DNA solutions were sheared 15 times through a 25-gauge hypodermic needle to reduce the molecular weight. DNA of low molecular weight is easier to make into films and avoids some of the problems with orientational artifacts. The DNA solutions were dialyzed exhaustively against $10^{-3}$ or $5 \times 10^{-4}$ *M* LiCl or $10^{-3}$ *M* $NH_4CH_3COO$ prepared with glass distilled water. d-AT:AT was concentrated by allowing several milliliters of 9 OD/ml d-AT:AT solution inside of a dialysis bag to evaporate to reduced volume. The resulting solution was exhaustively dialyzed against $10^{-3}$ *M* LiCl.

### Preparation of Films

The polymer films were formed on quartz plates 2 cm square. Using disposable micropipets, 50–150 $\mu$l of polymer solution were deposited in the center of the plate and spread with a glass rod to cover an area about 1 cm in diameter. The films were stirred with an up-and-down motion of the rod to avoid orienting the polymer in the plane of the plate. We found it easiest to form films when the polymer concentration was about 25 OD/ml. The salt content of the films was determined as previously described.[5] A mean molecular weight of 327 per residue and an extinction coefficient of $6.46 \times 10^3$ for *E. coli* DNA were assumed for that determination.

### Humidity Chambers

The relative humidity of the films was controlled by placing them inside sealed humidity chambers. The chambers have been described in greater detail elsewhere.[5] The chambers were used in a horizontal position, with the instrument light beam passing through in a vertical direction. The polymer film was deposited on the bottom quartz plate. The relative humidity inside the chambers was controlled by including a saturated solution of an appropriate salt. The salts and their corresponding relative humidities

were $Na_2HPO_4$, 95%; $NaC_2H_3O_2$, 92%; BaCl, 88%; KBr, 84%; $NaClO_3$, 75%; and $NaNO_2$, 66%. When the relative humidity was changed, films were allowed to equilibrate for at least 24 hr before CD measurements were taken. The chambers, once mounted in the instrument, could be rotated 360 degrees about the optical axis or tilted ±30 degrees.

## Modification of Instruments

CD measurements were made with a 6001 CD attachment to a Cary 60 spectropolarimeter. The instrument was modified so that the light beam and 6001 optical unit can be placed on a vertical axis, as well as in the normal horizontal mode. The modification is the same equipment described previously.[5] The instrument light beam is deflected into the vertically mounted optical unit by a front surface mirror placed in front of the Pockel's cell. Absorption measurements were done with a Bausch and Lomb spectronic 600 spectrophotometer. The spectrophotometer was placed in the vertical mode by bolting the instrument on a frame and tilting it 90 degrees on its end.

A special cell holder was used, which can hold a standard 1-cm pathlength cell in both the vertical and horizontal modes. Employing this cell, we checked the calibration of both orientation modes of the Cary 60 at 20°C with a solution of D-10 camphosulfonic acid (1 mg/ml). We also checked both modes for nonactive sample artifacts using a potassium dichromate solution (60 mg/l in $10^{-2}$ $M$ $H_2SO_4$) from 540–190 nm. The maximum artifact was less than 1 millidegree/OD of nonactive sample.

## Measurement of Films

Absorption spectra of the films were recorded from 350–220 nm before every CD measurement. The cells were positioned in the spectrophotometer so that the light beam impinged on approximately the same part of the film as in the Cary 60. The amount of polymer in the light path was determined from the absorption at 260 nm and the extinction coefficient of the polymer in solution. The absorption reading at 350 nm was subtracted from the 260-nm reading for this determination. DNA has no absorption at 350 nm. Any extinction in this region is due either to reflection from the various surfaces or to light scattering from the films, which is sometimes apparent at low humidities. The absorption baseline for an empty humidity chamber is flat. If scattering contributes to the 350-nm extinction however, there will be some error in the above procedure for determining absorption, since scattering extinction varies with wavelength. The extinction at 350 nm amounts to about 10% or 20% of the extinction at 260 nm, when the spectrophotometer is adjusted to read zero extinction at 350 nm with no chamber in the light path.

The CD baseline of the empty humidity chamber complete with salt boat and windows was recorded before each film was formed. The baselines of the three chambers used in these experiments are reproducible,

similar to each other, and relatively flat down to 220 nm. At 220 nm the baselines curve significantly and are not always reproducible. There may be errors in the reported CD spectra below 220 nm due to distorted baselines. These errors are small compared to the large ellipticities we observe in the lithium polynucleotide films, but they could be significant in the ammonium DNA spectra (see below). Since a film may be studied for periods of up to a week or more, the CD machine baseline may change in that time. When a film is measured several days after it is formed, a new baseline is run using another empty cell. All scans are made from 350 nm, and the pen position is adjusted to give a zero reading in that region. Circular dichroism is reported in units of $\epsilon_L - \epsilon_R$, where $\epsilon_L$ is the per residue extinction coefficient of left circularly polarized light in l/mol·cm.

## RESULTS AND DISCUSSION

### Ammonium DNA Films and A-Form Spectra

In order to provide a check on our present technique of making films as well as on past results[4,5] we measured the CD spectra of films of the ammonium acetate salt of *E. coli* DNA. Figure 1 shows the spectra of such a film at several relative humidities. The peak at 270 nm in the wet film shifts about 10 nm toward the blue and a negative minimum at 210 nm develops as the relative humidity is lowered. At low relative humidities a negative long wavelength band appears in the spectrum at around 290 nm. These results are in agreement with previous work[4,5] and have been interpreted as representing the CD spectrum of DNA in the A-form conformation.[6] In some films the initial negative minimum is quite large and the spectra are nearly conservative in the 240–300-nm region. This conservative spectrum is illustrated in Figure 2. The sign of the conservative spectral component is opposite to the sign of the conservative component of the B-form CD spectrum.

The most striking feature of the ammonium salt film spectra is the large magnitude of the 210-nm minimum relative to the 235–245-nm minimum. The increase in magnitude of the 210-nm minimum relative to the 235–245-nm minimum as the film goes into the A form also occurs in the previously reported spectra of the films of the sodium and ammonium salt of *E. coli* and calf thymus DNA,[4,5] and is probably the most characteristic feature of A-form DNA CD spectra, at least for these nucleic acids.

The A-form film DNA CD spectra have all the features of double-stranded RNA CD spectra,[7] except that the magnitudes of the positive 260-nm maximum are too small. The 260-nm maximum of double-stranded RNA of rice dwarf virus has a magnitude of about $\Delta\epsilon = 7$; our films have magnitudes for that maximum of about $\Delta\epsilon = 3$–4. Recently Ivanov et al.[8] measured the CD spectrum of calf thymus DNA in 78% ethanol and found it to have features of an A-form spectrum, including a magnitude of the 260-nm maximum of about $\Delta\epsilon = 9$ and a large negative 210-nm minimum.

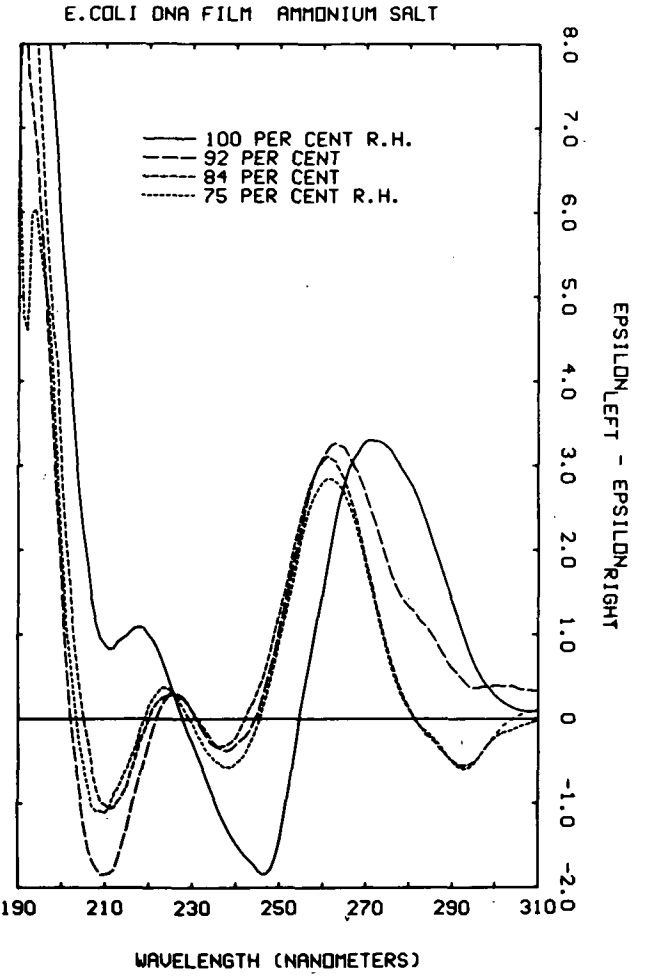



Fig. 1. Film of ammonium salt of *E. coli* DNA as function of relative humidity.

There are two possible explanations for the small magnitude of the 260-nm peaks of the film spectra. First, the transformation of DNA to the A form could be incomplete, leaving part of the DNA in another conformation. Alternatively, the DNA in the films could have a conformation that is in the family of A-form conformations, but that is somewhat different from that of double-stranded RNA of rice dwarf virus or DNA in 78% ethanol. Ivanov et al.[8] describe the various DNA conformations as belonging to two families: the B type and the A type. The two families are differentiated by the pucker of the ribose ring, which is C3-exo for the B family and C3-endo for the A family,[9] as well as by other well-known parameters such as base twist and tilt. Within families, various conformations are generated by changes in the winding angle and small changes in twist and tilt. Thus, B- and C-form DNA are within the B family, and A and

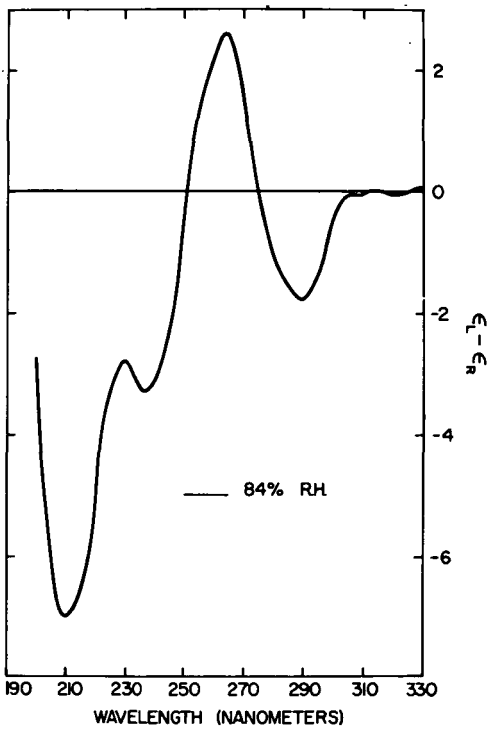



Fig. 2. Film of ammonium salt of *E. coli* DNA at 84% relative humidity showing conservative A-form CD spectrum.

A′ RNA are within the A family. Changes of polymer conformation within a family are influenced by changes in the salt in the polymer environment.[8,10] It is conceivable that our A-form film CD spectra represent the spectrum of an A-family DNA conformation, which is produced by the particular salt conditions of the films. Such a result would be analogous to the CD changes in DNA as it goes from the B conformation in dilute salt to the C conformation in high salt solutions. In the C conformation the magnitude of the long wavelength peak is much less than for the B conformation, while the short wavelength region of the CD spectrum is nearly unchanged.[4] It appears probable that there is not just one A-form DNA CD spectrum, but rather a whole family of them. There is the nonconservative spectrum seen in DNA in 78% ethanol, with a large 265-nm positive maximum and a weaker negative minimum at 245 nm. In double-stranded RNA of rice dwarf virus a negative band appears in the spectrum at 295 nm, and in some DNA film spectra that negative band is nearly as strong as the 260-nm positive peak, giving a conservative long-wavelength spectrum as in Figure 2.

## Lithium Films and Ψ-Form Spectra

As wet films (100% relative humidity) of the lithium salt of DNA are brought to relative humidities of about 92% or less, a dramatic change

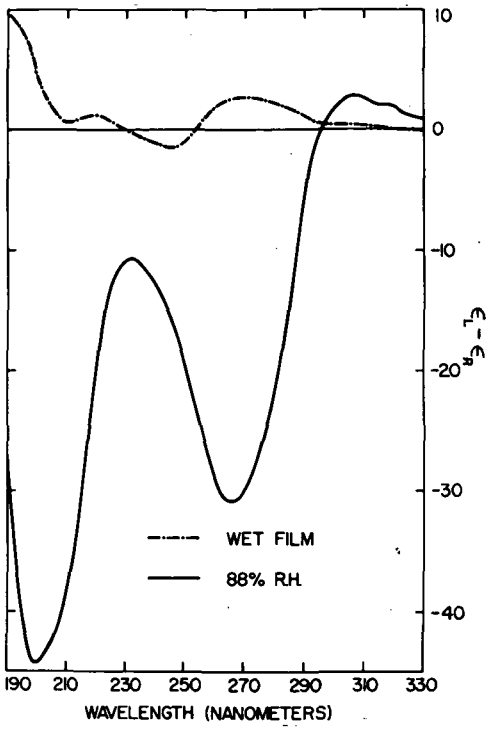



Fig. 3. Film of lithium salt of *E. coli* DNA at 100% and 85% relative humidity.

occurs in the CD spectrum. Figure 3 shows the CD spectra of a wet film of *E. coli* DNA, and of the same film equilibrated against 88% relative humidity (RH). The 100% RH CD spectrum resembles the solution spectrum for *E. coli* DNA. The 88% spectrum is nonconservative over the observable spectral range, with a large negative minimum at 265 nm, and another, larger negative minimum at 202 nm. The small positive peak at 305 nm is probably an artifact due to differential light scattering, since DNA does not absorb light in that region. This particular film, however, did not show the increased extinction at 350 nm upon dehydration to 88% RH, which implies light scattering by the film. Once the relative humidity is lowered to about 92%, further change in relative humidity does not influence the CD spectrum significantly. Figure 4 shows the CD spectra of one film studied at several relative humidities. The lithium salt concentration of both these films was between 2% and 4%.

Figure 5 shows the CD spectrum of a film of the lithium salt of d-AT:AT at 95% RH. The spectrum is similar in shape to the low-humidity spectra of the DNA films, but the magnitude of the d-AT:AT CD minima is much greater than for DNA. The CD minima in Figure 5 are at 270 and 210 nm.

To investigate the influence of salt concentration, we prepared *E. coli* DNA films with 0.5%, 1%, and 2% LiCl by weight. In each case negative

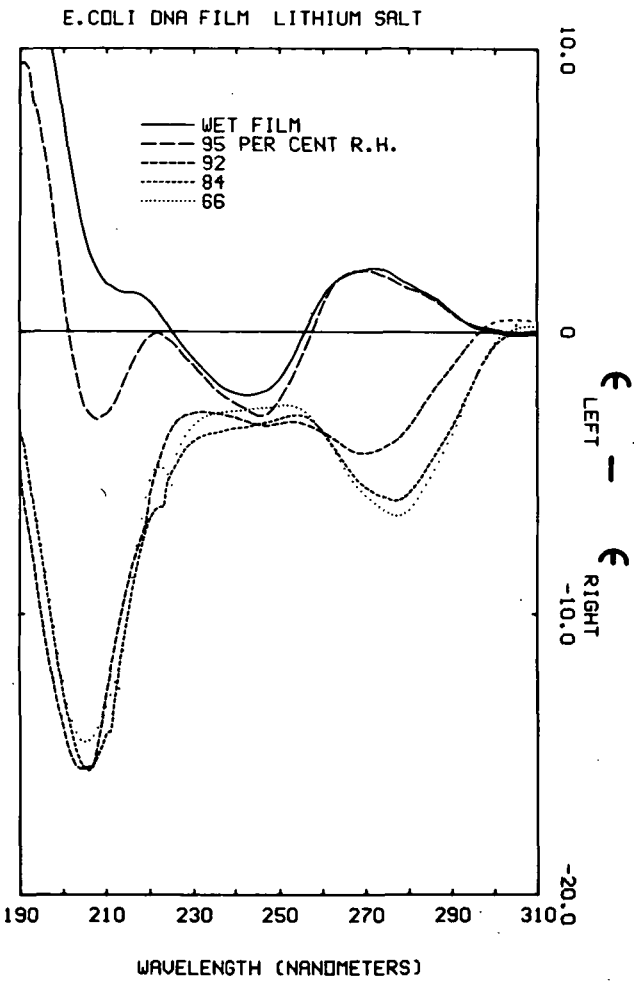



Fig. 4. Film of lithium salt of *E. coli* DNA as function of relative humidity.

nonconservative CD spectra with minima at about 265 and 205 nm were observed when the films were dehydrated to 92% RH. Similar results were obtained with films of calf thymus DNA.

The spectra in Figures 3–5 are typical of the film spectra we have obtained with the lithium salt of DNA or d-AT:AT. These results can be summarized by describing the spectra as negative and nonconservative, with a negative minimum between 265 and 270 nm, and another greater negative minimum between 200 and 210 nm. The magnitude of these nimima is not a distinctive feature of these spectra; it varies from film to film and must be influenced by the conditions of film preparation. The range of magnitudes of these minima is from $-5$ to $-300$. Films of d-AT:AT always give much larger minima than films of DNA. Similar spectra have been reported from solutions of DNA with high concentrations of neutral

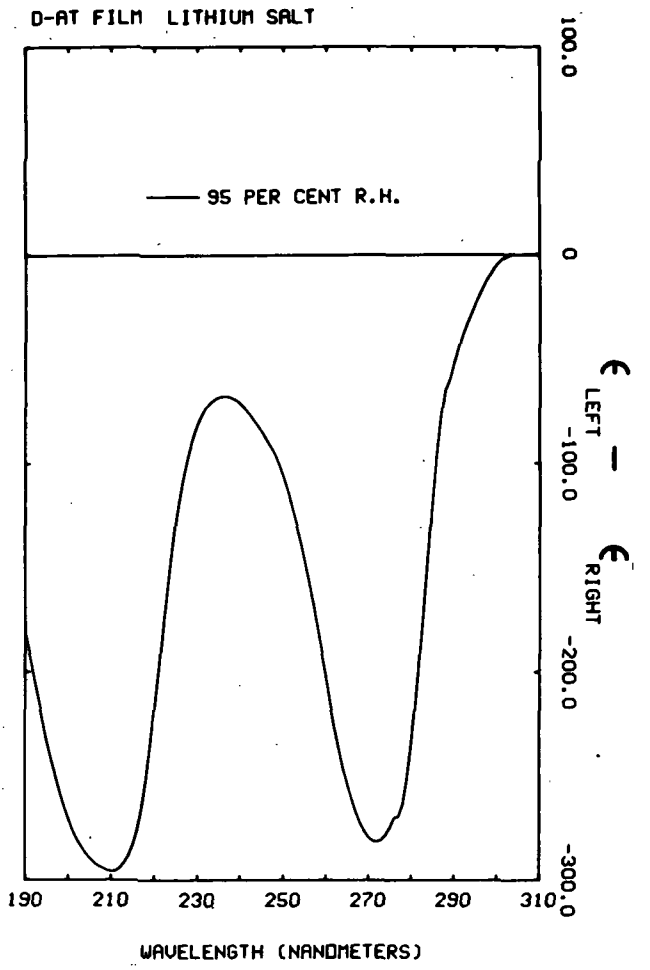



Fig. 5. Film of lithium salt of d-AT:AT at 95% relative humidity.

polymer.[11] Following the terminology of Jordon et al.[11] we will designate the above described spectral form as a Ψ-type CD spectrum.

To examine the possibility that the Ψ-form spectrum is an artifact resulting from orientation of the polymer on the quartz plates, several films were measured again after the chambers were rotated or tilted. The shape and magnitude of the film CD spectra were independent of rotation about the optical axis. The shape of the spectra also remained constant when the chambers were tilted, although sometimes the magnitude of the minima changed. The change of magnitude upon tilting can be explained by the fact that the light beam of the instrument passes through a different part of the film when the humidity chamber is tilted, and the films are not entirely uniform. We conclude that the Ψ spectrum is not an artifact due to orientation of the sample.

CD spectra with the same features that we have described for Ψ-form film spectra have been observed by Tunis-Scheider and Maestre[4] in some films of the ammonium acetate salt of calf thymus DNA. Ψ-type CD spectra have also been observed for a variety of reconstituted nucleohistone complexes and model compounds. ORD Cotton effects consistent with Ψ-type spectra have been reported in DNA complexed with poly-L-lysine.[12,13] Ψ-type CD spectra of DNA–polylysine complexes have also been observed by several authors.[14,15] The magnitude of the CD minima or Cotton effects is not constant; it varies from preparation to preparation.[13] Ψ spectra have also been reported from complexes of the lysine-rich histone fraction f-1 with DNA.[15–19] Again the magnitude of the minima varies from preparation to preparation, but most of the spectra exhibit minima with magnitudes between −5 and −35. The origin of the Ψ-type CD spectra in protein–DNA complexes has been attributed to changes in the secondary structure of the DNA molecule as it complexes with the protein fraction,[12,13] or to changes in the DNA secondary structure specifically induced by the protein component as the complexes aggregate into larger particles.[16–19]

It is reasonable to assume that the Ψ-form film and protein complex spectra have a similar origin. In that case the Ψ spectrum cannot be due to changes in DNA specifically induced by any protein. It is conceivable that the Ψ-type spectrum arises from interactions between bases of DNA that is in some new double-stranded conformation, much as the A- and B-form CD spectra are assumed to be produced.[20] The most prominent feature of the Ψ-type spectra is that they are overwhelmingly nonconservative throughout the entire observable spectral region. Theoretical treatment of polymer CD demonstrates that CD spectra that are nonconservative in a given spectral region must be due to interactions with transitions outside that region.[21] If Ψ spectra did reflect secondary DNA conformational changes, the CD spectra would be due to very strong interactions between the 260- and 210-nm absorption bands of DNA and the far uv transitions beyond 185 nm.

Neither the A- nor the B-form DNA CD spectra are nearly as nonconservative as the Ψ spectrum. Recent calculations have shown that it is possible to explain the A- and B-form spectra in a qualitative, approximate manner by interactions among nucleic acid base transitions in the near uv region.[22] If the Ψ-type spectrum repesents some new conformation of DNA, there would have to be much greater interaction between the near and far uv transitions in the Ψ form than in either the A or B geometries. That fact suggests that the hypothetical Ψ geometry would be considerably different from both the A and B structures.

There have been a variety of X-ray scattering studies on nucleohistone in gels and solution.[14,23–25] Although the resolution is poor, all the studies agree that the conformation of DNA in nucleohistone is very near to the B geometry. Some X-ray studies have suggested that the DNA might be supercoiled,[24,25] but measurements have shown that the CD spectrum of supercoiled DNA is very similar to that of the B form.[26] It is unlikely that

the Ψ-type CD spectrum is produced primarily by any change in the secondary or tertiary conformation of the DNA molecule. Interpretation of Ψ-form CD spectra from DNA protein complexes in terms of changed local DNA structure is probably erroneous.

The recent X-ray work of Suwalsky and Traub[27] does not deal with the structure of Ψ-form DNA since they made their DNA–protein complexes by direct addition, rather than by annealing through salt gradient dialysis. Previous CD studies indicate that DNA in direct addition complexes is in the B form.[15]

Gottesfeld et al.[28] have proposed that the Ψ-type spectrum might be an artifact resulting from differential light scattering from reconstituted nucleohistone complexes. Recently Dorman and Maestre have developed techniques to estimate the contribution of differential light scattering to measured CD.[29,30] Those techniques have been applied to complexes of DNA with f-1 histone.[31] The results demonstrate that while there is a contribution to f-1 DNA CD spectra due to differential light scattering, it is relatively small and cannot account for the observed features of the Ψ-type CD spectrum. Šponar and Frič[32] suggested on the basis of their observed dependence of the magnitude of the Ψ-type spectrum on the G–C content of DNA that the Ψ spectrum is due to a conformational change of the G and C bases and an interchain interaction between G bases. We have observed the strongest Ψ-type spectra in films of d-AT:AT however, so the role of guanine–guanine interactions in the production of Ψ-type CD spectra cannot be decisive.

Several authors[13,14] have proposed that the Ψ-type spectra of nucleohistone complexes results from liquid crystals.[33] If DNA molecules aggregate to form large, periodic, asymmetric structures, very large optical rotations could be expected. Tunis-Schneider and Maestre[4] suggested that DNA aggregation into quasicrystalline microdomains was responsible for the large CD values they observed in some of their films. Clearly the Ψ spectra result from some kind of aggregation; such spectra are found when DNA is highly concentrated in films or brought together in histone complexes. The appearance of Ψ spectra in f-1 histone complexes is dependent upon aggregation of the complexes.[14,16,19] Histone-complex spectra are also concentration dependent. Intense Ψ spectra are observed at lower histone to DNA ratios in $10^{-3}$ *M* DNA than in $10^{-4}$ *M* DNA.[16] Holzwarth and Holzwarth[34] have shown that liquid crystals can produce strong nonconservative CD spectra within the absorption bands of the crystal. Calculations by Holzwarth et al.[35] further indicate that long-range periodic asymmetric structure contributes to the observed CD spectrum of T2 phage. The liquid crystal theory also provides an explanation for the fact that the magnitudes, but not the shapes, of the Ψ spectra vary over a wide range. We believe the liquid crystal approach provides the most likely interpretation of these spectra.

The conditions under which our lithium polynucleotide films were prepared have been previously reported[4] to cause films to exhibit a different kind of CD spectrum than the one we report here. The previously observed

spectrum was identified with the C-form geometry of DNA.[36] The conditions under which lithium films exhibit the C-form spectrum, and the conditions under which the DNA aggregates to yield a Ψ-form spectrum, are presently not understood. However, it is known from X-ray fiber work that the salt content of the fiber as well as the relative humidity is critical to obtaining a C-form X-ray pattern.[36,37] Too much salt as well as too little can prevent fibers from assuming the C form. The salt content of our films has been only crudely estimated. It is possible that a careful study of films of various salt contents would define the conditions for the C- and Ψ-form CD spectra.

## CONCLUSIONS

The features we observed in the Ψ CD spectra of films of the lithium salt of DNA and d-AT:AT are similar to the features of some reported CD spectra of DNA–polylysine complexes and complexes of DNA with the f-1 fraction of calf thymus histone. It is reasonable to assume that the origin of these spectra is similar in both the films and the DNA–protein complexes. In that case the CD spectra observed in f-1 complexes are not the result of DNA conformational changes specifically induced by the protein fraction, as has been suggested by Fasman et al.[16–19] It is also unlikely that the Ψ spectra are the result of any DNA conformational change. Probably such spectra arise from long-range periodic asymmetric structures developing as DNA aggregates to form liquid crystals. The CD spectra of f-1 complexes are essentially an artifact and do not reveal information about the secondary structure of nucleohistone. Any CD changes that might occur as a result of DNA conformational changes as the polynucleotide complexes with the protein are masked by the much more intense Ψ spectrum.

CD spectra different from both DNA in solution and Ψ-type spectra have been reported for other protein–DNA complexes.[14,19,28,38–40] Whether these spectra represent specific conformational chnges in DNA, or whether they also represent CD artifacts, is presently unknown. It is evident however, that when unusually large or nonconservative CD spectra are observed from condensed or aggregated samples, one should suspect the possibility of artifacts due to liquid crystal phenomena.

This work was supported by NASA Grant NGR-05-003-460 and USPHS Grant AI-08427-Q4.

## References

1. J. T. Yang and T. Samejima, *Progress in Nucleic Acid Research and Molecular Biology*, J. N. Davidson and W. E. Cohn, Eds., Academic, New York, 1969, Vol. 9, p. 224.

2. C. R. Cantor and I. Tinoco, *Methods of Biochemical Analysis*, D. D. Glick, Ed. Wiley, New York, 1970, Vol. 18, p. 81.

3. J. Brahms and S. Brahms, *Fine Structure of Proteins and Nucleic Acids*, G. Fasman and S. Temasheff, Eds., Dekker, New York, 1970, Vol. 4, p. 191.

4. M. J. B. Tunis-Schneider and M. F. Maestre, *J. Mol. Biol.*, **52**, 521–541 (1970).

5. M. F. Maestre, *J. Mol. Biol.*, **52**, 543–556 (1970).
6. S. Arnott, S. Dover, and Wonacott, *Acta Cryst.*, **25**, 2192–2206 (1969).
7. T. Samejima, *J. Mol. Biol.*, **34**, 39–48 (1968).
8. V. I. Ivanov, L. E. Minchenkova, A. K. Schyolkina, and A. I. Poletayev, *Biopolymers*, **12**, 89–110 (1973).
9. S. Arnott and W. L. Hukins, *Biochem. Biophys. Res. Commun.* **47**, 1504–1509 (1972).
10. S. Arnott, W. Fuller, A. Hodgson, and I. Prutton, *Nature*, **220**, 561–564 (1968).
11. C. F. Jordan, L. S. Lerman, and J. H. Venable, *Nature New Biol.*, **236**, 67–70 (1972).
12. P. Cohen and C. Kidson, *J. Mol. Biol.*, **35**, 241–245 (1968).
13. J. T. Shapiro, M. Leng, and G. Felsenfeld, *Biochemistry*, **8**, 3219–3232 (1969).
14. M. Haynes, R. A. Garrett, and W. B. Gratzer, *Biochemistry*, **9**, 4410–4416 (1970).
15. D. Carroll, *Biochemistry*, **11**, 421–426 (1972).
16. G. D. Fasman, B. Schaffhausen, L. Goldsmith, and A. J. Adler, *Biochemistry*, **9**, 2814–2822 (1970).
17. A. J. Adler, B. Schaffhausen, T. A. Langan, and G. D. Fasman, *Biochemistry*, **10**, 909–913 (1971).
18. A. J. Adler and G. D. Fasman, *J. Phys. Chem.*, **75**, 1516–1526 (1971).
19. T. Y. Shih and G. D. Fasman, *Biochemistry*, **11**, 398–404 (1972).
20. C. W. Johnson and I. Tinoco, *Biopolymers*, **7**, 727–749 (1969).
21. I. Tinoco, *J. Chim. Phys.*, **65**, 91–97 (1968).
22. W. C. Brunner, Ph.D. dissertation, University of California, Berkeley, 1972.
23. S. Brahm and H. Ris, *J. Mol. Biol.*, **55**, 325–336 (1971).
24. S. Brahm, *J. Mol. Biol.*, **58**, 277–288 (1971).
25. J. F. Pardon, M. H. F. Wilkins, and B. M. Richards, *Nature*, **215**, 508–509 (1967).
26. M. F. Maestre and J. C. Wang, *Biopolymers*, **10**, 1021–1030 (1971).
27. M. Suwalsky and W. Traub, *Biopolymers*, **11**, 2223–2231 (1972).
28. J. M. Gottesfeld, M. Calvin, R. D. Cole, D. M. Igdaloff, V. Moses, and W. Vaughan, *Biochemistry*, **11**, 1422–1430 (1972).
29. B. P. Dorman and M. F. Maestre, *Proc. Nat. Acad. Sci. U.S.*, **70**, 255–259 (1973).
30. B. P. Dorman, J. E. Hearst, and M. F. Maestre, *Advances in Enzymology*, in press.
31. B. P. Dorman, and M. F. Maestre, manuscript in preparation.
32. J. Šponar and I. Frič, *Biopolymers*, **11**, 2317–2330 (1972).
33. C. Robinson, *Tetrahedron*, **13**, 219–234 (1961).
34. G. Holzwarth and N. A. W. Holzwarth, *J. Opt. Soc. Amer.*, **63**, 324–331 (1973).
35. G. Holzwarth, D. J. Gordon, J. E. McGinness, B. P. Dorman, and M. F. Maestre, manuscript in preparation.
36. D. A. Marvin, M. Spencer, M. H. F. Wilkins, and L. D. Hamilton, *J. Mol. Biol.*, **3**, 547–565 (1961).
37. A. Rupprecht and B. Forslind, *Biochim. Biophys. Acta*, **204**, 304–316 (1970).
38. V. I. Permogorov, V. G. Debabov, I. A. Sladkova, and B. A. Rebentish, *Biochim. Biophys. Acta*, **199**, 556–558 (1970).
39. D. Carroll, *Biochemistry*, **11**, 426–433 (1972).
40. T. Y. Shih and G. D. Fasman, *Biochemistry*, **10**, 1675–1683 (1971).

Received September 5, 1973
Revised November 12, 1973

D10

[Reprinted from the Journal of the American Chemical Society, 96, 4340 (1974).]
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner.

REPRODUCTION RESTRICTIONS OVERRIDDEN
NASA Scientific and Technical Information Facility

**Fluorescence Detected Circular Dichroism**

*Sir:*

Circular dichroism (CD) is a powerful technique for investigating molecular structure. The local structure near a chromophore in a macromolecule can be studied, if the CD of the chromophore can be isolated. We report here a new method, fluorescence detected circular dichroism (FDCD), which is capable of measuring the CD spectrum of a single chromophore in a macromolecule, provided the chromophore is fluorescent. Preliminary experimental results are given which show good agreement with theory for simple model systems.

The method is based on the usually valid assumption that the excitation spectrum of a fluorescent chromophore parallels its absorption spectrum.[1] This is generally true for systems in which energy transfer is negligible. In FDCD, the difference in absorption for left and right circularly polarized light is obtained by measuring the difference in fluorescence intensity for left and right circularly polarized excitation.[2] A schematic of the experimental arrangement is shown in Figure 1. The detection system of a Cary 60 spectropolarimeter with Model 6001 CD attachment is modified by placing a 5-cm diameter Hamamatsu R375 photomultiplier at 90° to the excitation beam. This photomultiplier is followed by a preamplifier and the resultant signal is processed by the Cary 60 electronics and displayed on the recorder chart. The pen position is digitized and smoothed in a manner described previously.[4] A colored glass filter is used to discriminate against scattered excitation light, and a standard quartz fluorescence cell is employed. In order to collect more fluorescence, the wall of the cell holder opposite the photomultiplier is made highly reflecting. Baselines are recorded by placing an optically inactive, fluorescent solution in the cell. For compounds with a large FDCD, sodium fluorescein proved adequate. However, for compounds with a small FDCD it is necessary for the baseline solution to approximate the fluorescent characteristics of the solution being measured. This is because the baseline seems to shift a small amount for different fluorescent species. The reasons for this

(1) G. Weber and F. W. J. Teale, *Trans. Faraday Soc.*, **54**, 640 (1958).
(2) Note that the polarization of the fluorescent light is not measured. This is a different method which gives information about the excited state of the chromophore.[3]
(3) J. Schlessinger and I. Z. Steinberg, *Proc. Nat. Acad. Sci. U. S.*, **69**, 769 (1972).
(4) A. D. Blum, O. C. Uhlenbeck, and I. Tinoco, Jr., *Biochemistry*, **11**, 3248 (1972).

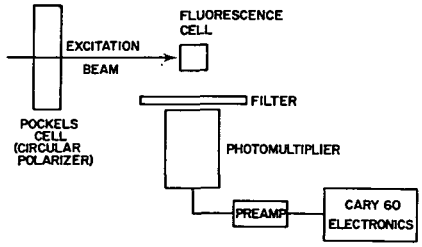



**Figure 1.** Experimental set-up for measuring FDCD.

shift are not yet understood, but thus far it has been a minor inconvenience.

The spectrometer measures $S = (F_R - F_L)/(F_R + F_L)$ as a function of excitation wavelength, where $F_R$ and $F_L$ are the fluorescence emission intensities for right and left circularly polarized exciting light, respectively. If one assumes that the quantum yield for fluorescence is independent of the wavelength and polarization of the exciting light, then for a solution containing only an optically active, fluorescent species, the CD is related to the FDCD by

$$\epsilon_L - \epsilon_R = \frac{-S(1 - 10^{-A})}{cd10^{-A}} \quad (1)$$

where $\epsilon_L$ and $\epsilon_R$ are the extinction coefficients for left and right circularly polarized light, $A$ is the absorbance due to the optically active molecule, $c$ is the concentration, $d$ is the pathlength, and $S$ is the measured value of the FDCD, which is equal to the ellipticity reading from the chart divided by 33.0 to give the correct units for $\epsilon_L - \epsilon_R$. This equation was tested by measuring the FDCD of morphine, *d*-10-camphor sulfonic acid, and L-tryptophan. Excellent agreement was obtained in each case. This agreement is strong evidence for the validity of our assumptions in these systems. The results for *d*-10-camphor sulfonic acid are shown in Figure 2. For this case, the absolute magnitude of $\epsilon_L - \epsilon_R$ at the 290 nm peak derived from the FDCD agrees with the previously published value to within 2%.[5]

The extraction of the CD spectrum from the FDCD spectrum is considerably more complicated when optically active, nonfluorescent species are present in addition to the fluorescent chromophore. Application of Beer's law results in the following equation for the CD of the fluorescent species

$$\epsilon_{FL} - \epsilon_{FR} = 2\epsilon_F\left(1 - \frac{2}{1 - \frac{A_L(S-1)(1-10^{-A_R})}{A_R(S+1)(1-10^{-A_L})}}\right) \quad (2)$$

where $\epsilon_{FL}$, $\epsilon_{FR}$, and $\epsilon_F$ are the extinction coefficients of the fluorescent chromophore for left circularly polarized, right circularly polarized, and unpolarized light, respectively; $A_L$ and $A_R$ are the absorbances for left and right circularly polarized light due to optically active material. The latter are obtained by measuring both the absorption and CD spectrum for the solution. This equation was verified by applying it to the FDCD spectra of mixtures containing approximately $10^{-5}$ *M* L-tryptophan and varying amounts of L-cystine in 0.025 *M* sodium phosphate buffer at pH 11. Under these conditions L-cystine is not fluorescent whereas L-tryptophan has a quantum yield of about 50%.[6] At the concentrations used, the CD of the mixture is reproduced by the sum of the CD of its components implying little or no interaction between the L-tryptophan and L-cystine. By measuring the FDCD, CD, and absorption spectra of the mixture, along with the absorption spectrum of L-tryptophan, it should be possible to extract the CD spectrum of L-tryptophan in the mixture. The FDCD baseline for this experiment was provided by scanning a solution of $1.37 \times 10^{-5}$ *M* DL-tryptophan. To test for artifacts, solutions with ten times the concentration of DL-tryptophan and mixtures of $1.37 \times 10^{-5}$ *M* DL-tryptophan plus up to $9.95 \times 10^{-5}$ *M* DL-cystine were also run. All these baselines were essentially coincident. Figure 3 shows a typical result derived from the FDCD of the active mixture. As predicted for little interaction between the components, the CD curve calculated from eq 2 is

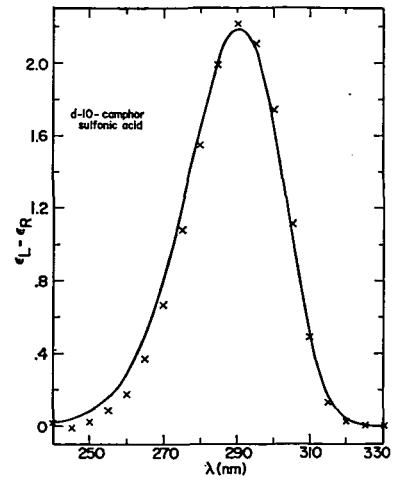



**Figure 2.** CD spectrum of *d*-10-camphor sulfonic acid: (——) CD spectrum, (×) points derived from FDCD and absorption spectrum using eq 1. Solution measured was $4.30 \times 10^{-3}$ *M*. Baseline for FDCD was $2 \times 10^{-5}$ *M* sodium fluorescein. Filter was Corning 4-72. Fluorescence cell had a 1-cm pathlength.

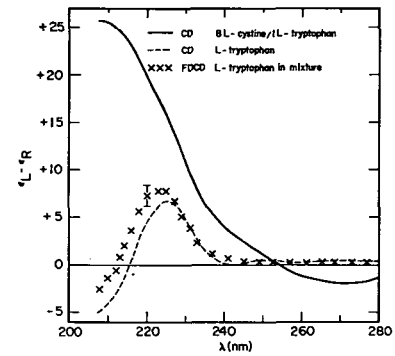



**Figure 3.** (----) CD spectrum of L-tryptophan, pH 11. (——) CD spectrum of $1.37 \times 10^{-5}$ *M* L-tryptophan plus $1.157 \times 10^{-4}$ *M* L-cystine, pH 11, per mole of L-tryptophan. (×) Points derived from FDCD, CD, and absorption spectra of $1.37 \times 10^{-5}$ *M* L-tryptophan plus $1.157 \times 10^{-4}$ *M* L-cystine using eq 2. Baseline used for FDCD was $1.37 \times 10^{-5}$ *M* DL-tryptophan, 0.025 *M* sodium phosphate, pH 11. Filter was Schott WG 295–2 mm. Fluorescence cell had a 0.3-cm pathlength. The error bar shows noise amplitude prior to final smoothing. Similar results were obtained with mixtures containing $1.37 \times 10^{-5}$ *M* L-tryptophan plus $5.79 \times 10^{-5}$ *M* L-cystine and $1.08 \times 10^{-6}$ *M* L-tryptophan plus $5.4 \times 10^{-6}$ *M* L-cystine.

(5) J. Y. Cassim and J. T. Yang, *Biochemistry*, **8**, 1947 (1969).

(6) R. W. Cowgill, *Biochem. Biophys. Acta*, **75**, 272 (1963).

very close to the CD curve for pure L-tryptophan. For comparison, the CD for the entire mixture is also shown.

These encouraging results on mixtures have led us to extend our experiments to macromolecules. Thus far we have measured the FDCD spectra of lysozyme and $t$-RNA$^{Phe}$. Here the fluorescent chromophores are the tryptophans and the anticodon loop Y base, respectively.[7,8] The signal-to-noise ratios in both spectra are several times larger than for the mixtures reported above. This is reasonable since the CD of free L-tryptophan is rather small relative to the CD of a residue in a protein or nucleic acid. To analyze these data, it will be necessary to obtain a value of $\epsilon_F$ for the chromophore in its macromolecular environment. There are several possible ways to accomplish this, including measurement of the corrected excitation spectra of the solutions.

The FDCD technique makes it possible to measure the CD of only the fluorescent chromophores in a macromolecule, provided there is little or no energy transfer. This greatly enhanced specificity for CD measurements is very difficult, if not impossible, to obtain by other methods.[9] Unique structural information on selected regions of biologically important macromolecules may thus be obtained.

**Acknowledgments.** We wish to thank Dr. E. Weitz, Dr. W. Hug, Dr. E. Pysh, Dr. K. Sauer, and Ms. C. Cech for several helpful discussions. This work was supported by Research Grant GM 10840 and AI-08427 from the National Institutes of Health and by the Atomic Energy Commission.

(7) S. S. Lehrer and G. D. Fasman, *J. Biol. Chem.*, **242**, 4644 (1967).
(8) U. L. Raj Bhandary, S. H. Chang, A. Stuart, R. D. Faulkner, R. M. Hoskinson, and H. G. Khorana, *Proc. Nat. Acad. Sci. U. S.*, 57, 751 (1967).
(9) S. Beychok, *Annu. Rev. Biochem.*, 37, 437 (1968).

**Douglas H. Turner,* Ignacio Tinoco, Jr.**
*Department of Chemistry and Chemical Biodynamics Laboratory*
*University of California*
*Berkeley, California 94720*

**Marcos Maestre**
*Space Sciences Laboratory, University of California*
*Berkeley, California 94720*

*Received March 8, 1974*

[Reprinted from Biochemistry, (1974) 13, 126.]
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner.

D10
D11

# Mie Scattering Contributions to the Optical Density and Circular Dichroism of T2 Bacteriophage†

G. Holzwarth,* David G. Gordon, James E. McGinness,‡ Burton P. Dorman,§ and Marcos F. Maestre

ABSTRACT: The contributions of Mie scattering to the absorption and circular dichroism (CD) spectra of T2 phage are calculated for a spherically symmetric model of the phage. The diameter of the model is 1000 Å. DNA occupies a central core of diameter 850 Å; this is covered by a protein coat of thickness 75 Å. The optical properties of the protein and DNA are taken from experimental absorption, CD, and optical rotatory dispersion (ORD) curves for disrupted phage, isolated DNA, and isolated protein ghosts. The absorption, scattering, total CD, and scattering contributions to the CD of the model particle are calculated. The calculation yields three conclusions: (1) Mie scattering is the predominant source for the turbidity of suspensions of intact T2 phage; (2) the signs and the shape of the Mie scattering contribution to the CD resemble those observed experimentally, but the observed amplitude of the 280-nm peak is much larger than that calculated and the observed long-wave tail is not found in the calculated curve; (3) the calculated CD curve of the model phage does not accurately yield the observed spectrum of the intact particle. This suggests that a DNA conformational change or long-range asymmetric packing of the DNA, rather than Mie scattering, are the dominant factors generating the observed differences between the CD spectra of intact and disrupted T2 phage.

The secondary structure and packing of DNA in bacteriophage T2 could be elucidated by optical spectroscopy if one understood the observed differences in both the circular dichroism (CD) and in the optical rotatory dispersion (ORD) of intact and disrupted bacteriophage T2 (Maestre and Tinoco, 1967; Maestre *et al.*, 1971). These differences have at least three possible sources: (1) Mie (isotropic particle) scattering due to the particulate nature of the phage suspension; (2) alteration of the optical activity of individual molecules through a change in secondary structure; (3) scattering contributions due to long-range chiral anisotropic packing of the DNA molecules in the phage head. How important is each of these mechanisms? The question can be approached experimentally, as by measurements made with and without the collection of scattered light (Dorman and Maestre, 1973), or by calculation. Our approach here is to estimate by calculation the magnitude of Mie scattering by the phage (first source listed above).

Mie scattering is a direct consequence of the dense packing of DNA in the phage head. The compact structure of the phage causes the refractive index inside the phage to be much greater than in the solvent. This generates large variations in refractive index at various points in the suspension; the refractive index variations lead, in turn, to light scattering effects on absorption and CD spectra. For the sake of the Mie calculation, the phage is modeled by a spherically symmetric particle having an isotropic DNA core and isotropic protein coat of dimensions comparable to those observed. Since Mie theory assumes spherical symmetry, the tail of the phage is deliberately ignored. However, the total mass of protein assumed for the coat includes the mass of the tail. The absorbance, turbidity, circular dichroism, and optical rotation of this model are calculated exactly by a method which has successfully explained particulate effects in the CD of red blood cell ghosts (Gordon and Holzwarth, 1971b) and aggregates of poly(L-glutamic acid) (Gordon, 1972). This method is based on Aden and Kerker's (1951) extension to spherical shells of Mie's (1908) method for evaluating scattering by simple spheres. Maxwell's equations are solved exactly, and no assumptions are made concerning the relative size of particle and wavelength. This is essential, since for ultraviolet (uv) spectroscopy of T2 the particle diameter and light wavelength are of comparable magnitude. Schneider (1971, 1973) has presented a similar method for evaluation of the optical properties of large particles. The calculated spectra are compared to spectra obtained for intact and disrupted phage, measured by a novel technique which allows essentially all of the scattered light to be collected (Dorman and Maestre, 1973; Dorman *et al.*, 1973).

## Experimental Section

T2 bacteriophage was chromatographed on a hydroxylapatite column and then dialyzed against a solution containing 0.5 M NaCl, 1 mM $MgSO_4$, and 1 mM Tris-HCl at pH 6.8, as previously described (Maestre *et al.*, 1971; Dorman and Maestre, 1973). Phages were disrupted by freeze–thawing. This procedure leads to release of DNA into the dialysis buffer. The protein coat and tail remain intact as a "ghost." Viral DNA was extracted with phenol followed by dialysis against 0.1 M NaCl in 10 mM sodium–phosphate buffer at pH 7.2.

† Contribution from the Department of Biophysics (G. H.) and the Department of Chemistry (D. J. G. and J. E. M.), University of Chicago, Chicago, Illinois 60637, and from the Department of Chemistry (B. P. D.), and the Space Sciences Laboratory (M. F. M.), University of California, Berkeley, California 94720. *Received September 10, 1973.* The research at the University of Chicago has been supported by the National Institutes of Neurological Diseases and Stroke through Grant NS-07286. The work at the University of California was supported by NASA Grant 05-003-020 and by National Institutes of Health Grants AI-08427-04 and GM-11180. G. H. is grateful for the support of Research Career Development Award GM-15050 from the National Institutes of Health. D. J. G. is pleased to acknowledge support by U. S. Public Health Service Training Grant No. HD-00001 from the National Institute of Child Health and Human Development.

‡ Present address: Department of Chemistry, University of California, La Jolla, Calif. 92037.

§ Present address: Department of Biology, Yale University, New Haven, Conn. 06520.



REPRODUCTION RESTRICTIONS OVERR[illegible]N
NASA Scientific and Technical Information Facil[illegible]

Optical density curves were measured in a standard cuvet by a normal Cary 14 spectrophotometer. Although this instrument undoubtedly collects part of the scattered light, we assume that the fraction of scattered light collected is small, and we designate spectra thus obtained as $A_T$, representing total (absorptive and scattering) optical density. To obtain true absorption spectra free of the effects of light scattering, the sample was placed in a special fluorscat optical cell (Dorman *et al.*, 1973) in Cary 1462 scattered transmission accessory. Absorbance curves thus obtained are labeled $A_A$. The contribution of scattering to the measured optical density is labeled $A_S$ and is equal to $A_T - A_A$.

Circular dichroism curves with and without collection of scattered light were similarly obtained. The circular dichroism measured in the usual way with a standard cuvet, in a Cary 6001 or 6003 CD instrument, is designated $\theta_T$ on the assumption that no scattered light is accepted by the detector. CD measurements in which scattered light and transmitted light were collected, yielding $\theta_A$, the absorptive part of the circular dichroism, were obtained using a fluorscat cell in a Cary 6001 instrument. The scattering contribution to circular dichroism, $\theta_S$, is given by $\theta_T - \theta_A$. Occasionally, measurements of $\theta_T$ were obtained by placing a large-area photomultiplier tube directly adjacent to the sample cuvet. Data similar to those of the fluorscat technique were obtained.

Optical rotation data, $\Phi$, measured with a standard cuvet in a Cary 60 spectropolarimeter, were taken from previously published work (Maestre and Tinoco, 1967); no correction was made for scattering.

## Calculations

The calculations are performed for a spherically symmetric model of the virus, as shown in Figure 1. The total radius of the particle, $R$, is taken as 500 Å, in accord with the average radius of the phage head (Brenner *et al.*, 1959; Cummings and Kozloff, 1962). (A recent review by Cummings *et al.* (1970) suggests that 400 Å would be a better average radius; this radius yields results differing only slightly from those for the 500-Å radius, as discussed later.) The particle is composed of a protein coat of thickness $\Delta$ and a DNA core of radius $R - \Delta$. The value of $\Delta$ was chosen as 75 Å; electron microscopic evidence suggests that the protein layer is 60–80 Å thick (Kellenberger *et al.*, 1968). There are thus three regions of space for the calculation: 1 = DNA core, 2 = protein coat, and 3 = external solvent. For each region of space we need to know the complex refractive index for left and right circularly polarized light; these are designated $m_{1L}$, $m_{1R}$, $m_{2L}$, $m_{2R}$, etc. After the refractive indices are specified, one can obtain, by a rigorous, straightforward calculation, the total optical density $A_T$, the scattering $A_S$, and the absorbance $A_A$, for a suspension of particles. The total optical density $A_T$ is calculated by considering the detector to accept only transmitted light and light which is scattered at zero angle. The scattering contribution to the optical density, labeled $A_S$, is obtained by integrating the scattered radiation for all scattering angles. Finally, the absorptive contribution $A_A$ is evaluated simply as $A_T - A_S$. The quantity $A_A$ corresponds to a measurement of absorbance in an instrument which collects the transmitted light and all scattered light ($4\pi$ sr).

In a similar fashion, the Mie calculation yields the total, scattering, and absorptive circular dichroism $\theta_T$, $\theta_S$, and $\theta_A = \theta_T - \theta_S$. The quantity $\theta_T$ corresponds to a CD measurement by an instrument whose detector collects only transmitted light and light scattered directly forward; $\theta_A$ corresponds to a CD measurement in which transmitted light and all scattered light reach the detector.

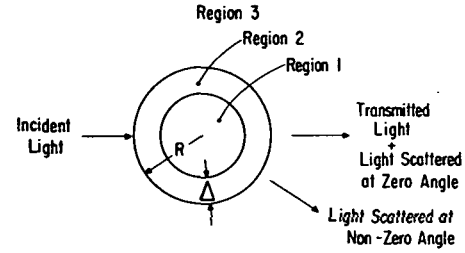



FIGURE 1: The spherically symmetric structure used in the calculations to model T2 phage. The outer particle radius $R$ is 500 Å. Region 1, the core, of radius $R - \Delta$, is occupied by DNA and water; region 2, of thickness $\Delta = 75$ Å, is occupied by protein and water. Region 3 is occupied by water alone. For light incident from the left, a detector placed an infinite distance away toward the right would see only transmitted light and light scattered at zero angle. By contrast, light scattered at nonzero angles will also be accepted if a large area detector is placed close behind the scatterer or if the scatterer is placed in a fluorscat cell.

The method used to evaluate the complex refractive indices at a given wavelength $\lambda$ is as follows. For region 1, the DNA core, we use a mean dielectric constant containing contributions from both water and DNA. The contribution of DNA is estimated by evaluating the Kronig–Kramers transform $K$ of the absorption spectrum of isolated, B-form T2 DNA. We use the measured absorption spectrum between 190 and 350 nm and extrapolate the data to zero at 140 nm. To the term $K$ we add the Kronig–Kramers transform of a lumped absorption band at 120 nm, giving a term $Q\lambda^2/(\lambda^2 - \lambda_0^2)$. The size $Q$ of the second band is adjusted to yield $dn/dc$ equal to 0.188 ml/g at 435.8 nm, as observed (Reichmann *et al.*, 1954). For any desired DNA concentration $c$ in grams per milliliter one has

$$\epsilon_1(c) = 1 + (\epsilon_{H_2O} - 1)(1 - c\bar{v}) + c\bar{v}(K + Q\lambda^2/[\lambda^2 - \lambda_0^2]) \quad (1)$$

Here $\bar{v}$ is the partial specific volume of DNA, taken as 0.55 $cm^3/g$ (Hearst, 1962). The value of $c$ is fixed by the core size and the known molecular weight of the T2 DNA, $1.2 \times 10^8$ daltons (Stent, 1963). The mean real refractive index of region 1, $n_1$, is just $\epsilon_1^{1/2}$.

The complex refractive indices for left and right circularly polarized light are then evaluated, in the manner previously described (Gordon and Holzwarth, 1971b; Gordon, 1972), as follows

$$m_{1L} = \bar{n}_1 - i\lambda 2.303A_1/4\pi NV_1 + \lambda(\Phi_1 - i\theta_1)/360NV_1 \quad (2)$$

$$m_{1R} = \bar{n}_1 - i\lambda 2.303A_1/4\pi NV_1 - \lambda(\Phi_1 - i\theta_1)/360NV_1 \quad (3)$$

Here $V_1$ is the volume of the core. $A_1$, $\Phi_1$, and $\theta_1$ are the absorbance (base 10), ORD (degrees), and CD (degrees) of a T2 DNA solution for a 1-cm path and a standard concentration, which we took as $10^{-4}$ M in nucleotide (or $N = 1.9 \times 10^{11}$ phage/$cm^3$). The absorbance, ORD, and CD data were all measured by normal methods in standard cuvets, since these solutions do not scatter light significantly.

The refractive indices for region 2 are calculated similarly. The mean dielectric constant of pure protein is approximated by the Kronig–Kramers transform of a Gaussian absorption band at 190 nm (maximum molar absorptivity $10^4$ $M^{-1}$ $cm^{-1}$;

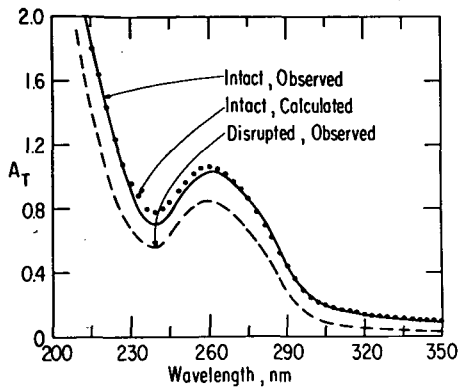



FIGURE 2: The observed optical density spectrum $A_T$ of intact (—) and disrupted (- - -) T2 phage, and the corresponding spectrum $A_T$ calculated for intact phage (· · ·). The data include the effects of light lost by scattering and by absorption. Phage concentration: $10^{-4}$ M in nucleotides ($1.87 \times 10^{11}$ phages/ml). Observed data were measured with Cary 14 in a standard 1-cm cuvet.

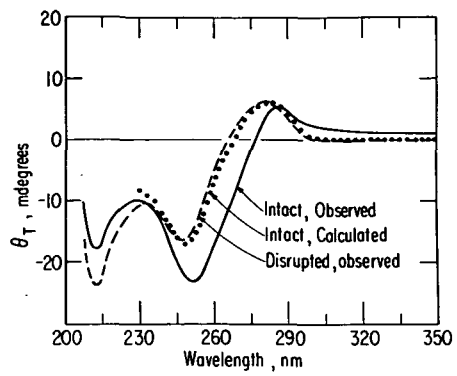



FIGURE 4: Circular dichroism of intact and disrupted T2 phage as measured by normal methods: (—) CD ($\theta_T$) for intact phage as measured with a Cary 6003 using standard cuvet; (· · ·) $\theta_T$ calculated for our model of the intact phage; (- - -) CD measured for disrupted phage using a standard cuvet in a Cary 6001 instrument: path length, 1 cm; concentration, $10^{-4}$ M in nucleotides. The CD is given in millidegrees of ellipticity.

bandwidth 12 nm) and an additional band at 120 nm with intensity $Q'$. The value of $Q'$ is adjusted to yield $dn/dc$ = 0.194 ml/g at 435.8 nm; this value of $dn/dc$ is an average of published values for a variety of proteins (Doty and Edsall, 1959). The dielectric constant for a given shell geometry, assuming a protein mass of $9 \times 10^7$ daltons (Stent, 1963), is evaluated by the analog of eq 1. The partial specific volume of protein was taken as 0.72 cm³/g (Tanford, 1961).

The complex refractive indices for left and right circularly polarized light are then evaluated by eq 2 and 3, but with $A_1$ replaced by $A_2$ (the absorbance of T2 ghosts), $\Phi_1$ replaced by $\Phi_2$ (T2 ghost rotation), and $\theta_1$ replaced by $\theta_2$, the ghost CD. The ghost CD was obtained by subtracting the CD of DNA from the CD of disrupted phage as measured with fluorscat optics. Both $A_2$ and $\theta_2$ are measured by methods which eliminate the effects of scattering; $\Phi_2$ is not so corrected, but the error introduced should be small in the spectral region of interest. The refractive index $m_3$ is that of the solvent, water, so $m_{3R} = m_{3L}$. The values of $m_3$ are interpolated from data in the International Critical Tables.

The real mean refractive indices at several wavelengths of the DNA core, protein coat, and water, in that order, are as follows: λ 350, $\bar{n}_1 = 1.473$, $\bar{n}_2 = 1.496$, $\bar{n}_3 = 1.349$; λ 300, 1.504, 1.516, 1.359; λ 250, 1.509, 1.555, 1.377; λ 220, 1.569, 1.609, 1.398. These refractive indices correspond to a DNA concentration of 0.623 g/ml and a protein concentration of 0.736 g/ml as dictated by the particle geometry.

With the above specifications for $R$, $\Delta$, $m_{1L}$, $m_{1R}$, $m_{2L}$, $m_{2R}$, and $m_3$, the optical spectra $A_T$, $A_S$, $A_A$, $\theta_T$, $\theta_S$, $\theta_A$, and $\Phi$ of the model T2 phage suspension are readily calculated by Mie theory (Gordon, 1972; eq 5, 10, and 13–21). These results may then be compared to the respective observed spectra of T2 phage.

## Results

The experimental data to be explained, and the results of the Mie calculations, are presented in Figures 2–5. Figure 2 shows the experimental total optical density (absorbance plus scattering), $A_T$, for intact phage and for disrupted phage. There is substantially greater optical density observable in

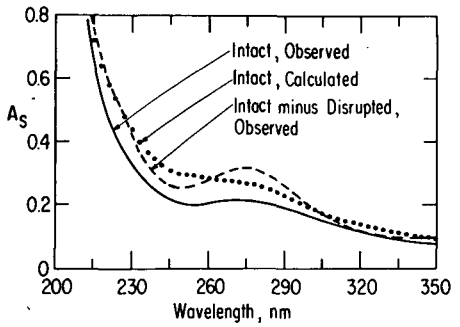



FIGURE 3: The contribution of scattering to the optical density of T2 phage: (—) intact phage, optical density measured with Cary 14 in standard cuvet, *minus* optical density measured in fluorscat cell with Cary 1462 scattered light collection accessory; (· · ·) scattering contribution $A_S$ calculated for intact phage; (- - -) optical density of intact phage measured with standard cuvet in Cary 14, minus absorption of disrupted phage measured with fluorscat cell in a Cary 1462. The dashed curve may reflect several other factors in addition to scattering. Path length, 1 cm; concentration, $10^{-4}$ M in nucleotides.

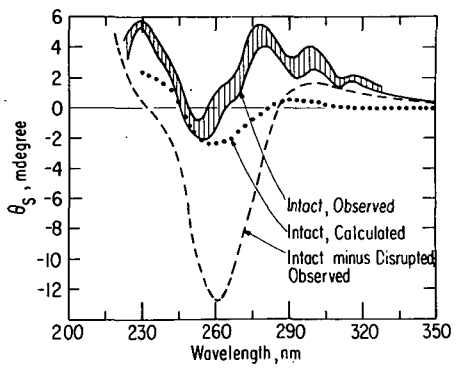



FIGURE 5: Scattering contributions to the circular dichroism of T2 phage. The two solid lines are two separate experiments evaluating $\theta_S$ for intact phage. Each shows the difference between CD measured for intact phage with standard cuvet *minus* the CD measured for intact phage with Cary 6001 using fluorscat optical cell. Dots show the calculated scattering contribution $\theta_S$ to the CD of our model of the intact phage. The dashed curve gives the difference between CD measured for intact phage in a standard cuvet with Cary 6003 minus the CD measured for disrupted phage in a Cary 6001 using a standard cuvet and a close detector. The disrupted phage CD spectrum has been shown to be independent of the detection geometry (Dorman and Maestre, 1973). Path length, 1 cm; phage concentration $10^{-4}$ M in nucleotides. The units of $\theta$ are millidegrees of ellipticity.

the intact phage than in disrupted phage for all wavelengths shown. The intact phage spectrum shows a pronounced tail for $\lambda > 320$, which arises from scattering.

The calculated values of $A_T$ for our model of the intact phage are also shown in Figure 2. The calculations are seen to fit the observed $A_T$ quite closely; the long-wave tail and the detailed shape of the enhancement in optical density, both of which are due to scattering, are nicely predicted by our model.

The role of scattering in the absorption spectra can also be studied experimentally using the fluorescent scattering technique (Dorman and Maestre, 1973). The value of $A_S$ for intact phage, evaluated by subtracting the measured absorbance of T2 suspension in a fluorscat cell from that for a normal cuvet, is shown in Figure 3. The experimental curve shows a substantial tail for $\lambda > 320$, a level region from 290 to 245 nm, and then rises sharply. The calculated values of $A_S$ are seen to agree closely in magnitude and shape with the observed curve. This result supports the view that for intact T2 phage, the scattering contribution to optical density is largely a consequence of dense packing of DNA inside the phage head. By contrast, the difference between $A_T$ for intact phage and $A_A$ for disrupted phage, also given in Figure 3, shows a small peak at 275 nm and a modest dip at 245 nm. These features, which are not as prominent in the experimental and calculated $A_S$ curves, originate in a 2-nm red-shift of the absorption maximum of the DNA in intact phage compared to that of the disrupted particle. In addition, the enhanced amplitude of the difference between $A_T$ for intact phage and $A_A$ for disrupted phage reflects an 11% absorbance increase (hyperchromic shift) of the DNA in the intact phage relative to the solution spectrum of B-DNA. This hyperchromism is discussed further below.

The corresponding experimental and calculated total CD curves (absorptive plus scattering), $\theta_T$, are shown in Figure 4. Consider first the experimental data for intact phage. These show a pronounced long-wave tail, a modest, positive peak at 288 nm, and a large, negative trough with a minimum at 250 nm. The curve is distinctly nonconservative. By contrast, the disrupted phage, with B-form DNA, shows no scattering tail. The 280-nm peak is larger, while the 248-nm trough is shallower than that for the intact phage, leading to a more conservative CD curve for the 260-nm absorption region. The calculated CD for the intact phage, also shown in Figure 4, closely resembles the CD of the disrupted phage. Neither the long-wave CD tail of the intact phage, nor the nonconservative shape of the CD curve, is reproduced by the calculation. We conclude that the observed difference between $\theta_T$ for intact and disrupted phage cannot originate solely from the particulate, dense packing of DNA in the phage head.

In Figure 5 are shown the corresponding experimental and calculated curves of $\theta_S$, the scattering contribution to the circular dichroism. The experimental value of $\theta_S$ for intact phage is evaluated by subtracting the absorptive CD, measured by the fluorscat method, from the CD measured with normal optics. There is no change in particle structure between the two measurements. The experimental $\theta_S$ curve shows a 4–5-mdeg positive peak near 280 nm, a trough, $-1$ to $-2$ mdeg, near 255 nm, and then a positive peak again at wavelengths shorter than 245 nm. The calculated $\theta_S$ curve shows a positive peak, $+0.6$ mdeg, at 290 nm, a negative trough, $-2.4$ mdeg, at 257 nm, and positive CD for $\lambda < 245$ nm. The observed and calculated curves thus have the same basic shape, although the magnitude of the positive scattering peak for $\lambda > 270$ differs in the two cases. This suggests that a substantial part of the experimental $\theta_S$ may originate in the same physical properties which give rise to Mie scattering from T2 phage. It is useful to contrast this result with the observed difference between experimental $\theta_T$ curves for intact and disrupted phage, also shown in Figure 5. This curve has a small positive peak at 300 nm, a very large negative trough, $-13$ mdeg, at 260 nm, and positive CD for $\lambda < 230$. This curve is not accounted for by the Mie scattering calculation or by the observed $\theta_S$.

How sensitive are the calculated curves to the particular parameters chosen to represent the phage? We have performed similar calculations with the refractive index of protein and/or DNA incremented by 0.1. The calculated values of $A_T$ and $A_S$ obtained systematically exceed the observed by a substantial amount, so we believe our original refractive indices to be accurate. The amplitude of the $\theta_S$ curve is increased by larger $n$, but the $\theta_T$ and $\theta_S$ curves are not qualitatively altered and our conclusions on the fit between theory and experiment would remain unchanged.

We have also varied the particle radius to 400 and 600 Å, and the protein coat thickness to 50 and 100 Å. The qualitative results obtained are consistent with the increased or decreased refractive indices which automatically follow from eq 1 when a fixed amount of DNA or protein is confined to a larger or smaller compartment volume. Again no significant improvement in the fit of calculated CD curves to the observed could be achieved without the calculated $A_S$ grossly exceeding that observed.

We note that all of our Mie calculations have been performed using complex refractive indices derived from disrupted phage, for which the DNA is in the B conformation. It would be desirable to use molecular properties appropriate to the conformation of the DNA in the phage. Indeed, data for the CD and absorption curves of the DNA and protein of intact phage, free from the effects of scattering, packing, and absorption flattening, are a goal of the present study. However, we elected not to use as input data for eq 2 and 3 CD and absorption curves measured on intact phage with fluorscat optics because we do not have available as input for the Mie calculations corresponding experimental ORD curves for intact phage which are free of scattering effects. Since the calculated scattering contribution $\theta_S$ is of the same shape as the input ORD curve for the DNA, an undistorted input ORD is essential to the calculation of $\theta_S$. However, the resemblance of $\theta_S$ to the input ORD (Gordon, 1972) allows the reader to readily gauge the shape of the calculated $\theta_S$ curve which would result from assuming different DNA conformation. In particular, if C-form data were used as input to the Mie calculation, the positive peak in the calculated $\theta_S$ near 290 nm would largely disappear, giving, for $\lambda > 280$ nm, a poorer fit between the experimental and the calculated $\theta_S$ curves.

## Discussion

*Mie Scattering.* We are now in a position to answer several questions about the absorption and CD spectra of T2 and about the structural interpretation of these spectra. First, we assert that the scattering contribution to optical density, $A_S$, can be fully accounted for by the dense particulate nature of T2 phage. It is not necessary to invoke internal structure (anisotropy) of the DNA in the phage; such structure might be expected to yield enhanced scattering (Tinker, 1972). However, we cannot with confidence turn the argument around to assert that our results support a nonstructured packing of DNA in the phage, since the scattering from

spinach chloroplasts, which are known to be highly structured, can be largely explained by Mie calculations for an isotropic model (Bryant *et al.*, 1969). The difference in optical density $A_T$ for intact and disrupted phage (the dashed curve in Figure 3) appears to originate largely in Mie-type isotropic particle scattering, but additional factors, which are discussed below, must also be involved.

Turning next to circular dichroism, we note in Figure 4 that the calculated CD curve does not fit the CD of intact phage. Instead, it resembles the CD of disrupted phage, *i.e.* it resembles closely the data used as input to the calculation. Thus Mie scattering, which was invoked in the introductory statement as one of three possible sources to explain the difference between the CD of intact and disrupted phage, cannot explain the effects entirely, although its contribution is nonnegligible.

Is it possible that our neglect of the phage tail has led to a gross underestimate of the CD contributions arising from the dense, particulate nature of the phage? Experiments show that the phage ghost, which retains the tail, exhibits negligible CD between 245 and 350 nm (B. P. Dorman and M. F. Maestre, manuscript in preparation). The tail thus cannot play a major role in the CD of the intact phage.

*Secondary Structure vs. Anisotropic Packing.* Let us turn from the primary subject of this manuscript, Mie scattering, to consider the other factors which could play a significant role in the CD of intact phage. In so doing, we enter an area of suggestive rather than definitive experimental data; reasonable and prudent persons, not excluding the authors, may differ in their conclusions. The recently reported crystallization of T4 phage (Speyer and Khairallah, 1973) suggests that X-ray diffraction analysis may soon provide information which will dispel our present uncertainty about the secondary conformation of DNA inside the phage.

We consider first three experimental observations supporting the viewpoint, first adopted by Maestre and Tinoco (1967), that secondary structural changes play a major role in the CD changes accompanying phage disruption.

(A) Similarities between the CD spectra of C-form films (Tunis-Schneider and Maestre, 1970) and the fluorscat CD spectra of intact phage suggest that DNA in the phage is in C conformation (Dorman and Maestre, 1973). Both spectra show very little CD for $\lambda > 280$ nm, in contrast to B-form DNA, which shows a large positive band in this spectral region. Moreover, the intrinsic CD of B- and C-form DNA have been shown to originate in large terms of opposite sign which nearly cancel (Johnson and Tinoco, 1969). Thus it can be argued that small changes in helix geometry, such as an alteration of the base–tilt angle, could lead to large changes in the CD. The persuasiveness of the C-DNA assignment is moderated by the observation that the major 250-nm trough, $-7$ mdeg deep in the CD of intact T2 phage, is not quite so prominent ($-3.5$ mdeg) in the C-form spectra of calf thymus DNA films (Dorman and Maestre, 1973; Tunis-Schneider and Maestre, 1970). However, this difference quite probably is due to the extensive (70%) glucosylation of T2 DNA, which has been shown to be the dominant cause of enhancement of the DNA CD troughs at 210 and 245 nm (Usatyi and Shlyakhtenko, 1973).

(B) Spectrophotometric data demonstrate that DNA inside T2 and T6 phage possesses an $OD_{260}$ 33% greater than the same DNA in the B-form conformation (Rubenstein *et al.*, 1961; Bonhoeffer and Schachman, 1960). In more recent work, Tikchonenko *et al.* (1966) have shown spectrophotometrically that this change originates not only in changes in scattering, but also in 11% hypochromicity upon release of the phage DNA. These results are corroborated by our own work with the fluorscat technique, which yields a total optical density loss of 32+% upon T2 disruption and an 11% drop between the fluorscat spectra of intact and disrupted T2 (B. P. Dorman and M. F. Maestre, manuscript in preparation). We interpret these absorption data as evidence for a DNA conformational change.

(C) The magnetic CD spectrum of intact T2 phage exhibits no sign of a long-wavelength tail but does show spectral properties which can in some respects be mimicked by the hyperchromic DNA conformation produced through partial heat denaturation *in vitro* (Maestre *et al.*, 1971). This is evidence that conformational change occurs upon phage disruption.

The evidence favoring CD contributions from anisotropic packing of DNA in the intact phage is as follows. As we have noted above, the calculated Mie scattering contribution to the CD of intact phage, Figure 5, fails to explain the large, positive, long-wave tail in the observed $\theta_S$ and $\theta_T$ curves. This discrepancy probably does not originate in the assumption of sphericity, since the absorption curve is nicely fitted. Conformational change cannot explain the discrepancy for nonabsorptive regions ($\lambda > 320$). Moreover, conformational change is an unlikely source of the discrepancy for $270 < \lambda < 320$, since this would require that the DNA inside the phage has ORD more positive than B-form DNA for $\lambda > 280$ nm. The available data, admittedly not corrected for scattering, suggest instead that the DNA in the phage has negative ORD in this region (Maestre and Tinoco, 1967). Correction of these data for Mie scattering can be estimated from our calculations; the net ORD would probably remain negative for $\lambda > 280$, unlike the ORD data for the B form.

It is interesting to note parenthetically that, in contrast to the present case, anisotropic structure need not always be invoked to explain the CD of light scattering specimens. For example, the spectral differences between CD curves for intact and dispersed red blood cell ghosts *are* explicable by a Mie calculation similar to the one performed here for T2 (Gordon and Holzwarth, 1971b). The red blood cell ghost calculation assumes an isotropic, spherically symmetric model although one knows that the membrane must be optically anisotropic due to lipid and perhaps protein orientation in the plane of the membrane. One may speculate that anisotropy is of little influence on the red cell ghost spectra for two reasons. (1) The large radius of curvature (35,000 Å) and small membrane thickness (70 Å) combine to yield a particle which behaves optically like two independent 70-Å thick sheets. This thickness is much less than the wavelength. (2) The intrinsic CD of the membrane proteins is large compared to their absorbance, whereas the linear dichroism of the membranes is probably small. A calculation of the effects of anisotropy on scattering by 300-Å diameter, hollow phospholipid vesicles (Tinker, 1972) suggests a somewhat contrary view, but the ratio of thickness to diameter is much larger for the vesicles than for the red cell ghost.

What sorts of optical effects can anisotropy generate in CD spectra? It has been shown by several groups (Saeva and Wysocki, 1971; Saeva, 1972; Schrader and Korte, 1972; Sackmann and Voss, 1972; Dudley *et al.*, 1972; Chabay, 1972) that cholesteric liquid crystals exhibit relatively enormous CD and ORD bands centered approximately at the frequency of each of their absorption bands when studied by light propagating parallel to the cholesteric helix axis. (These bands occur in addition to the well-known optically active reflection band

(de Vries, 1951) which appears, generally in nonabsorptive spectral regions, when the light wavelength in the medium equals the helix pitch.) It is found that the CD band shapes generally resemble those of the absorption bands, but with a "hook," often generating CD of opposite sign, at their long- or short-wave edge (Chabay, 1972; Holzwarth *et al.*, 1973). This hook arises from the birefringence dispersion generated by the absorption band itself (Holzwarth and Holzwarth, 1973).

The CD effects of a randomly oriented collection of small liquid crystallites, such as could perhaps occur in T2 suspension, are at present not known; the experiments and theory for the liquid crystal CD in absorptive spectral regions have, so far, been confined to normal incidence on single crystals. However, recent studies of the CD of randomly oriented spinach chloroplasts and of polylysine–DNA complexes show instructive similarities to the CD of simple liquid crystals, as follows. Philipson and Sauer (1973) have shown experimentally that spinach chloroplasts exhibit very large CD effects in the 680-nm region of chlorophyll absorption. The CD bands possess tails extending into nonabsorptive spectral regions. The large CD effects are not eliminated when scattered light is collected in the CD measurement; instead, the bands change sign. However, when the supramolecular order (grana structure) of the chloroplast is broken down by sonication, the CD diminishes in amplitude by a factor of 10. Philipson and Sauer (1973) suggest that the large-amplitude CD of the intact chloroplast originates primarily in the ordered arrangement of chromophores into a grana structure rather than in the optical activity of individual chlorophyll molecules.

Similarly, studies of compact poly-L-lysine–DNA complexes (Cohen and Kidson, 1968; Shapiro *et al.*, 1969; Haynes *et al.*, 1970) show large nonconservative CD bands with maximum amplitude near 260 nm. The shape of the CD bands is strikingly similar to that of intact T2, yet X-ray diffraction studies indicate that the particles, which are comparable in size to T2, contain B-form DNA (Haynes *et al.*, 1970). Felsenfeld and coworkers (Shapiro *et al.*, 1969) and Gratzer and coworkers (Haynes *et al.*, 1970) suggest, as one possible explanation of their CD data, that the DNA in the particles might possess an ordered anisotropic supramolecular, perhaps liquid crystalline, arrangement.

The CD spectra of suspensions of chloroplasts and of polylysine–DNA complexes thus share three features with the CD curves of liquid crystals: (1) CD amplitudes apparently much larger than those intrinsic to their molecular constituents; (2) CD extending to nonabsorptive spectral regions; (3) CD band shapes closely resembling the absorption spectrum but with a hook of opposite sign.

Applicability of the liquid crystal results to T2 bacteriophage is not as convincing, since, for T2, the amplitude of the CD spectrum is only slightly greater in the intact particle than in the disrupted phage (item 1 above). However, strong scattering is observed well into the visible (item 2). This scattering cannot originate in the intrinsic CD of any particular DNA secondary structure or, as we now know, in Mie scattering. Finally, the difference CD spectrum between intact and disrupted phage (Figure 5, dashed line) bears a strong similarity to a DNA absorption curve, but with a hook at longer wavelengths (item 3 above).

## Summary

Mie scattering calculations for a model of T2 bacteriophage demonstrate three points. (1) The contributions of scattering to the optical density of T2 phage suspensions are correctly predicted by Mie theory. This suggests that the dense, particulate nature of the phage is the predominant source of scattering contributions to the OD. (2) The contributions of Mie scattering to CD of intact T2 phage are calculated to be +0.6 mdeg at 290 nm and −2.4 mdeg at 257 nm for a 1-cm path through a suspension $10^{-4}$ M in nucleotides. The corresponding experimental values of the scattering are 2.5–3.5 mdeg at 290 nm and 0 to −2 mdeg at 257 nm. The calculated CD spectral shape resembles the observed scattering curve broadly. However, it fails to explain the experimental scattering amplitude at 280 nm (4–5 mdeg) or the extent of the observed long-wave tail. (3) The calculated CD for intact phage, using as input data the spectra of disrupted phage, does not describe accurately the observed CD spectrum of the intact particle. The difference is believed to arise from a combination of conformational change and from supramolecular ordering. It is noted that the difference CD curve is similar in shape to CD curves observed at the absorption bands of liquid crystals.

## Acknowledgments

We are grateful to K. D. Philipson and K. Sauer for a preprint of their provocative work on chloroplast CD.

## References

Aden, A. L., and Kerker, M. (1951), *J. Appl. Phys. 22*, 1242.
Bonhoeffer, F., and Schachman, H. K. (1960), *Biochem. Biophys. Res. Commun. 2*, 366.
Brenner, S., Streisinger, G., Horne, R. W., Champe, S. P., Barnett, L., Benzer, S., and Rees, M. W. (1959), *J. Mol. Biol. 1*, 281.
Bryant, F. D., Seiber, B. A., and Latimer, P. (1969), *Arch. Biochem. Biophys. 135*, 97.
Chabay, I. (1972), *Chem. Phys. Lett. 17*, 283.
Cohen, P., and Kidson, C. (1968), *J. Mol. Biol. 35*, 241.
Cummings, D. J., Couse, N. L., and Forrest, G. L. (1970), *Advan. Virus Res. 16*, 1.
Cummings, D. J., and Kozloff, L. M. (1962), *J. Mol. Biol. 5*, 50.
Day, L. A., and Hoppensteadt, F. (1972), *Biopolymers 11*, 2131.
de Vries, H. (1951), *Acta Crystallogr. 4*, 219.
Dorman, B. P., Hearst, J. E., and Maestre, M. F. (1973), *Methods Enzymol. 27D*, 767.
Dorman, B. P., and Maestre, M. F. (1973), *Proc. Nat. Acad. Sci. U. S. 70*, 255.
Doty, P., and Edsall, J. (1959), *Advan. Protein Chem. 6*, 55.
Dudley, R. J., Mason, S. F., and Peacock, R. D. (1972), *J. Chem. Soc., Chem. Commun.*, 1084.
Gordon, D. J. (1972), *Biochemistry 11*, 413.
Gordon, D. J., and Holzwarth, G. (1971b), *Proc. Nat. Acad. Sci. U. S. 68*, 2365.
Haynes, M., Garrett, R. A., and Gratzer, W. B. (1970), *Biochemistry 9*, 4410.
Hearst, J. E. (1962), *J. Mol. Biol. 4*, 415.
Holzwarth, G., Chabay, I., and Holzwarth, N. A. W. (1973), *J. Chem. Phys. 58*, 4816.
Holzwarth, G., and Holzwarth, N. A. W. (1973), *J. Opt. Soc. Amer. 63*, 324.
Johnson, C., and Tinoco, I., Jr. (1969), *Biopolymers 7*, 727.
Kellenberger, E., Eiserling, F. A., and Boy de la Tour, E. (1968), *J. Ultrastruct. Res. 21*, 335.

Maestre, M. F., Gray, D. M., and Cook, R. B. (1971), *Biopolymers 10*, 2537.
Maestre, M. F., and Tinoco, I., Jr. (1967), *J. Mol. Biol. 23*, 323.
Mie, G. (1908), *Ann. Phys. (Leipzig) 25*, 377.
Philipson, K. D., and Sauer, K. (1973), *Biochemistry 12*, 3454.
Reichmann, M. E., Rice, S. A., Thomas, C. A., and Doty, P. (1954), *J. Amer. Chem. Soc. 76*, 3047.
Rubenstein, I., Thomas, C. A., Jr., and Hershey, A. D. (1961), *Proc. Nat. Acad. Sci. U. S. 47*, 1113.
Sackmann, E., and Voss, J. (1972), *Chem. Phys. Lett. 14*, 528.
Saeva, F. D. (1972), *J. Amer. Chem. Soc. 94*, 5135.
Saeva, F. D., and Wysocki, J. J. (1971), *J. Amer. Chem. Soc. 93*, 5928.
Schneider, A. S. (1971), *Chem. Phys. Lett. 8*, 604.
Schneider, A. S. (1973), *Methods Enzymol. 27D*, 751.
Schrader, B., and Korte, E. H. (1972), *Angew. Chem. 84*, 218; *Angew. Chem., Int. Ed. Engl. 11*, 226.
Shapiro, J. T., Leng, M., and Felsenfeld, G. (1969), *Biochemistry 8*, 3219.
Speyer, J. F. and Khairallah, L. H. (1973), *J. Mol. Biol. 76*, 415.
Stent, G. (1963), Molecular Biology of Bacterial Viruses, San Francisco, Calif., W. H. Freeman.
Tanford, C. (1961), The Physical Chemistry of Macromolecules, New York, N. Y., Wiley.
Tikchonenko, T. I., Dobrov, E. N., Velikodvorskaya, G. A., and Kisseleva, N. P. (1966), *J. Mol. Biol. 18*, 58.
Tinker, D. O. (1972), *Chem. Phys. Lipids 8*, 230.
Tunis-Schneider, M. J. B., and Maestre, M. F. (1970), *J. Mol. Biol. 52*, 521.
Usatyi, A. F., and Shlyakhtenko, L. S. (1973), *Biopolymers 12*, 45.

*Acta Astronautica.* Vol. 1, pp. 71–79. Pergamon Press 1974. Printed in the U.S.A.

A74-33730

# The impact of space electric field research on atmospheric studies

F. S. MOZER

Physics Department and Space Sciences Laboratory, University of California, Berkeley, California 94720

(*Received* 6 *September* 1973)

**Abstract**—Space measurements of electric fields have provided instrumentation for measuring atmospheric parameters and a better basis for understanding the electrical coupling between the magnetosphere and the atmosphere. Applications of an incoherent scatter radar (developed for ionospheric electric field research) to the measurement of atmospheric winds and turbulence and of Langmuir double probes (also developed for space research) for measurement of atmospheric electric fields are described. The increased knowledge of magnetospheric electric fields has focused attention on the electrical coupling between the magnetosphere and the atmosphere with conclusions that should considerably modify previous physical concepts in both domains.

## 1. Introduction

THE FIRST PROBLEM in discussing the impact of aerospace technology on atmospheric research is to define the upper altitude limit of the atmosphere. For electric field research, this limit is particularly important because of the vast amount of research on electric fields in the ionosphere and because of the different fundamental physical mechanisms operating in the ionosphere and at lower altitudes. By way of introducing this problem, the physics of electric fields in these two domains will first be reviewed.

The ionospheric medium is electrically coupled to the magnetosphere along magnetic field lines which are generally considered to be sufficiently good conductors that electric fields parallel to the magnetic field lines are small. Thus, potential differences imposed on the magnetosphere through its interaction with the solar wind (Axford, 1969; Gonzalez and Mozer, 1973) are transported along magnetic field lines to the ionosphere to produce ionospheric potential differences that are typically the order of 50 kV. There is considerable evidence that the solar wind interaction producing these potentials occurs through the process of magnetic field reconnection (Dungey, 1961) in which interplanetary magnetic field lines become connected to terrestrial field lines at the dayside boundary between the terrestrial and interplanetary media. The flow of the interplanetary magnetic field and plasma past the earth stretches these connected terrestrial field lines into an elongated tail containing magnetic energy that has been transferred to the magnetosphere from the solar wind. The release of this magnetic energy accelerates plasma to high energies, thereby producing the auroral and radiation belt particles. The interplanetary electric field associated with the motion of the solar wind through the magnetized interplanetary medium, is impressed on the magnetosphere along interplanetary magnetic field lines that become connected to ter-

restrial magnetic field lines, and thus, electric potentials are imposed on the magnetosphere through its interaction with the solar wind.

The electric field in the lower atmosphere probably results from vertical currents flowing through the resistive atmosphere to discharge the earth which is charged by lightning strokes. In the simplest picture, the earth-ionosphere cavity is represented as a leaky, spherical capacitor that is charged by thunderstorm activity and discharged by currents through the leaky dielectric. In this model the earth and ionosphere are assumed to be equipotentials that are at a potential difference of about 350 kV (Dolezalek, 1971).

Because the physics of electric fields in both the ionosphere and atmosphere is rather well understood and because such questions have been reviewed extensively in the recent past (see Axford, 1969, and the series of articles in *Pageoph* **100**, 1, 1971), these topics will not be covered further in the present discussion. Instead, an emphasis will be placed on the application of technologies developed for space electric field research to measurements at altitudes below about 80 km, and to the knowledge gained from space research on the electrical coupling between the atmosphere below 80 km and the ionosphere. Other possible topics, such as atmospheric electricity on other planets or the generation of electromagnetic waves in the magnetosphere by lightning strokes, will not be discussed in this brief summary.

## 2. Applications of space electric field instrumentation to atmospheric research

### (a) *Incoherent scatter radar*

The many types of instruments used for measuring electric fields in space have been reviewed in a recent article (Mozer, 1973). An example of the application of one such instrument to atmospheric research is the incoherent scatter radar at Jicamarca, Peru. The purpose of this instrument is to transmit 50 MHz electromagnetic pulses into the ionosphere and to measure the intensity and spectral distribution of return signals obtained from scattering in the ionospheric plasma. The Doppler shift of the return signal yields the line of sight velocity of the medium, from which the electric field, $\bar{E}$, may be obtained at altitudes above the collision dominated region from

$$\bar{E} = -\bar{v} \times \bar{B} \tag{1}$$

where $\bar{v}$ is the measured velocity and $\bar{B}$ is the ionospheric magnetic field. It has been recently shown that this instrument is able to measure signals scattered by variations of the index of refraction at altitudes between about 15 and 80 km (Woodman and Guillen, 1972). An example of the return signal strength during one such measurement is given in Fig. 1, in which signals over the altitude ranges of 15–35 and 60–85 km are significantly above the instrument noise level. A possible mechanism for the enhanced fluctuations of the dielectric constant at these altitudes is atmospheric turbulence in the presence of temperature (15–35 km) or density (68–85 km) gradients that carry regions of smaller dielectric constant into regions of enhancement (Booker and Gordon, 1957). Since the Doppler shift of

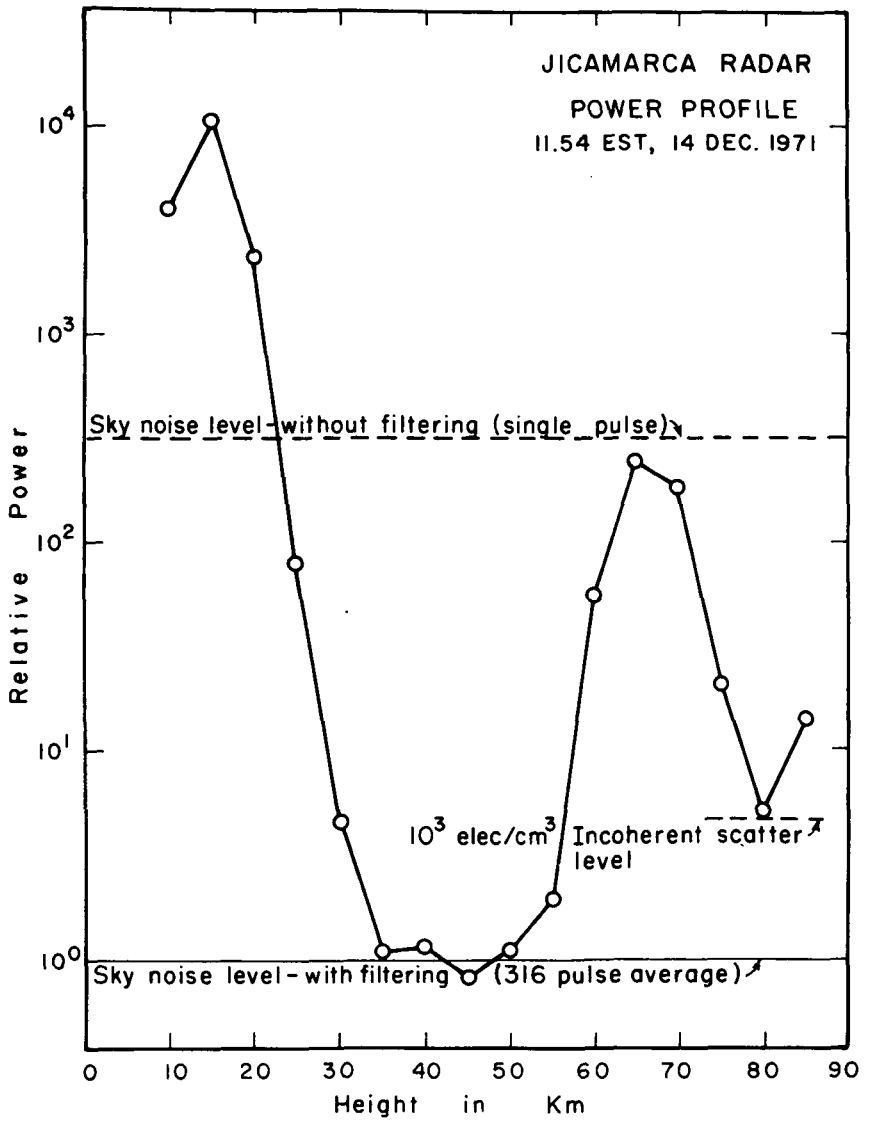



Fig. 1. Return signal strength as a function of altitude for electromagnetic signals back-scattered by atmospheric irregularities.

these return signals gives the wind velocity at these altitudes, significant information on turbulence and motions in an altitude region that is inaccessible to other techniques has been obtained by an instrument originally designed to study such parameters in and beyond the ionosphere.

(b) *Double Langmuir probe detectors*

Atmospheric electric fields have been measured at altitudes below about 30 km by use of field mills. Such devices measure the electric charge on a conducting

plate on which the atmospheric electric field terminates. This charge is related to the electric field by

$$\nabla \cdot \bar{E} = 4\pi\rho \tag{2}$$

where $\rho$ is the charge density on the plate. In practice, the charge is usually caused to flow on and off the plate by alternately exposing and grounding it, and the measured a.c. current is then proportional to the electric field present just outside the instrument. This field is simply related to the field that would be present in the absence of the instrument if the instrument does not greatly perturb the medium in its vicinity. This situation occurs in near vacuum or in the atmosphere but not in the ionospheric plasma, because the electric field near the instrument in a plasma results from the potential imposed on the instrument by the condition that the net thermal plasma current to it be zero in equilibrium.

With the assumptions that the plasma contains a single species of ions; that these ions and the electrons are at the same temperature, $T$; that the conductor is at rest in the plasma; and that the conductor is large compared to the Debye length; this potential may be computed by considering that the conductor acquires a net negative charge until it reaches a potential at which the current of electrons, energetic enough to overcome the potential and strike the conductor, is just equal, in magnitude, to the sum of the ion current; the current, $I$, to the measuring electronics; and the current, $I_p$, arising from non-thermal particles, photoemission, and all otherwise neglected sources. This current balance equation for an isotropic plasma in thermal equilibrium is

$$I + I_p + An \exp\left(\frac{kT}{2\pi m_i}\right)^{1/2} = An \exp\left(\frac{kT}{2\pi m_e}\right)^{1/2} \exp\left(\frac{eV}{kT}\right) \tag{3}$$

where $V$ is the potential of the conducting surface with respect to the nearby plasma, $m_i$ and $m_e$ are the ion and electron masses, $k$ is Boltzmann's constant, $n$ is the density of the plasma, and $A$ is the surface area of the conducting body. From Eq. (3), the potential that a floating body (one that is isolated such that $I = 0$) reaches in such a plasma is

$$V = \frac{kT}{e} \ln\left[\left(\frac{m_e}{m_i}\right)^{1/2} + \frac{I_p}{An \exp\left(\frac{kT}{2\pi m_e}\right)^{1/2}}\right] \tag{4}$$

The value of $V$ in the ionosphere, where $I_p$ is negligible, is typically about $-1$ V. Because of this potential, field mills cannot measure ionospheric electric fields, and a class of instruments, called Langmuir double probes, has been developed for space research. These instruments function through measurement of the potential difference between a pair of such separated conductors and division of this potential difference by the separation distance to obtain the electric field. The problem arising from the relatively large potential of Eq. (4) is eliminated in this measurement since this potential appears on both conductors of the pair and cancels when the potential difference between the conductors is obtained.

The double probe electric field instrument operates well in the ionosphere provided the potential measuring electronics has a sufficiently high input impedance that the current flowing through the electronics and back through the medium

does not cause the initial electric field to become distorted. The double probe instrument is superior to the field mill for upper atmospheric electric field measurements because of its greater sensitivity but may not function as well near the ground where the fields are too large to be easily accommodated in the dynamic range of the double probe. By use of such an instrument on approximately 100 balloon flights, nearly 1000 hours of atmospheric electric field measurements near an altitude of 30 km have been made at many latitudes (Mozer and Serlin, 1969; Mozer, 1971; Mozer and Manka, 1971). Some of these measurements will be discussed in the following section.

### 3. Relationships between magnetospheric and atmospheric electric fields

The classical leaky capacitor model of atmospheric electricity assumes that the ionosphere is an electrical equipotential (Dolezalek, 1971). It is the purpose of this section to show that this model is incorrect and to suggest that the electrical coupling between the atmosphere and magnetosphere may be significant to the physics in both domains. The experimental data used in this discussion come from Langmuir double probe measurements of electric fields on balloons.

One example of such data is given in Fig. 2, along with measurements of deviations of the surface magnetic field measured with a ground based magnetometer. Because the magnetometer responds primarily to ionospheric currents and because of the close correlation between the balloon electric field and ground magnetic field data, it is clear that the balloon borne detector measured electric fields whose origin was in the ionosphere. The physical mechanisms allowing the electric field produced at an altitude above 100 km to map to an altitude of 30 km where it dominates other fields, will be investigated within a framework in which the atmosphere is thought of as a medium having a scalar conductivity, $\sigma(z)$, that depends on the altitude, $z$. Electric fields are imposed on the upper boundary of this medium (at say, 100 km) by magnetospheric processes and on the lower boundary by weather associated phenomena. The question of the electric fields at any point within the medium is then answered by solution of a boundary value problem involving Maxwell's equations.

To solve the boundary value equations it is necessary to know the electrical conductivity of the atmosphere as a function of altitude. Many experiments (Sagalyn and Fitzgerald, 1965) have shown that the conductivity variation with altitude may be represented as

$$\sigma(z) = \sigma_0 \exp(z/h) \tag{5}$$

where $h$, the scale height of the conductivity variation, is approximately independent of $z$ and is about 6 km. The steady-state equations whose solutions are desired are

$$\nabla \cdot \bar{E} = 0 \tag{6}$$

$$\nabla \cdot \bar{J} = 0 \tag{7}$$

and

$$\bar{J} = \sigma \bar{E} \tag{8}$$

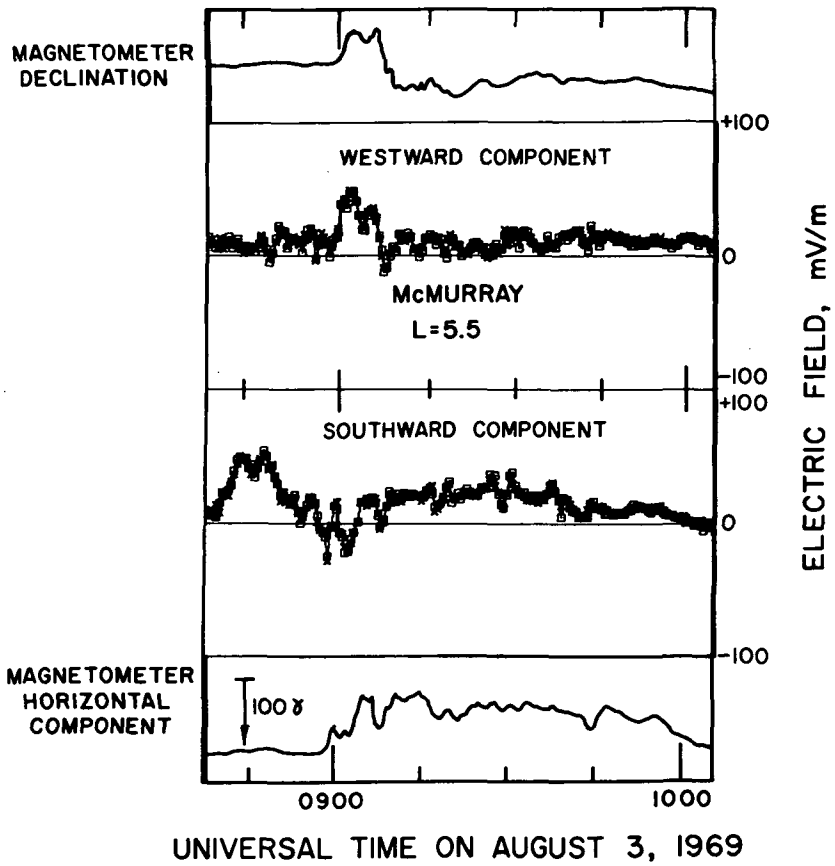



Fig. 2. Balloon horizontal electric field measurements and ground observations of magnetic field variations caused by ionospheric currents.

Satisfying Eq. (6) by assuming

$$\bar{E} = -\nabla V \tag{9}$$

the combination of Eqs. (5), (7), (8), and (9) yields

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} + \frac{1}{h}\frac{\partial V}{\partial z} = 0 \tag{10}$$

as the boundary value equation whose solution gives the electric field at balloon altitudes for particular boundary conditions.

To investigate how horizontal ionospheric electric fields map to balloon altitudes, it will be assumed that $E_x = E_1$, and $E_y = 0$ at an altitude $z_0$ ($\approx 100$ km) and that $E_x(z = 0) = 0$. At other altitudes the horizontal electric field is given by the solution to Eq. (10) as

$$E_x(z) = E_1 \frac{1 - \exp(-z/h)}{1 - \exp(-z_0/h)} \tag{11}$$

Since $z/h \sim 6$ at balloon altitudes, large spatial scale horizontal ionospheric electric fields map to balloon altitudes essentially without attenuation. Thus, ionospheric potential differences appear with essentially full strength to altitudes of 10–20 km and are only decreased by a factor of $e = 2.7$ at an altitude of about 6 km.

Typical large scale ionospheric potential differences are 50 kV, and they may occasionally be a factor of 10 larger (Gonzalez and Mozer, 1973). Thus, virtually the entire lower atmosphere is efficiently coupled to electric fields of magnetospheric origin which produce potentials that are comparable to the typically 350 kV potential resulting from weather processes. This coupling and its effect on atmospheric electricity should be one of the central topics of atmospheric research in coming years.

It is also likely that electric fields of atmospheric origin produce observable effects in the magnetosphere. To study this question, Park and Dejnakarintra

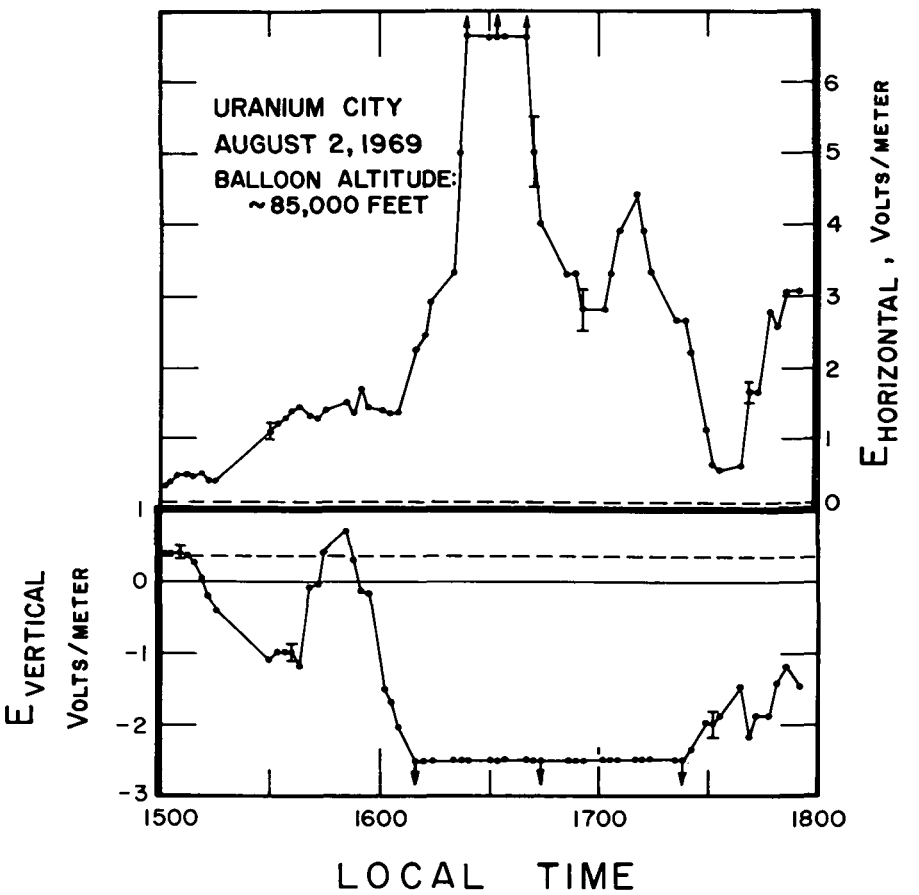



Fig. 3. Example of large electric fields measured on a balloon located above an afternoon thunderstorm.

(1972) have solved Eq. (10) for the atmosphere and its analog for non-scalar conductivities in the ionosphere to map thunderstorm electric fields from the lower atmosphere to the magnetosphere. They have found that large thunderstorms can produce equatorial electric fields in the magnetosphere having magnitudes that are 1–10% of those produced by magnetospheric processes. Such fields may be responsible for formation of localized electron density enhancements that are called magnetospheric ducts. An important objective of magnetospheric physics should be the experimental study of the mapping of such fields into the magnetosphere and their importance to magnetospheric dynamics.

An example of large thunderstorm electric fields observed on balloons is given in Fig. 3. The horizontal dashed lines of this figure indicate typical values of the field strengths observed at balloon altitudes and it is readily concluded that extremely large changes took place during the event of interest. The vertical electric field was inverted from its normally downward direction during most of the three-hour interval illustrated in Fig. 3 and the horizontal field was more than 100 times larger than the typical value observed during fair weather. The event occurred during the late afternoon in a region in which cumulonimbus clouds and thunderstorms were observed by several weather stations.

Figure 3 is of interest to problems of atmospheric electricity since it offers strong support for the classical view that thunderstorm activity provides a generator that drives global atmospheric electrical processes. It also illustrates the applicability and importance of balloon double probe electric field measurements to problems of atmospheric electricity. Thus, the instrument conceived to study space electricity has made and should continue to make significant contributions to problems of atmospheric electricity.

*Acknowledgments*—The balloon electric field experiments described in this paper were supported by NASA under contracts NAS 9-9502 and NAS 5-10362, by the National Science Foundation under grants GA-1317, GA-11259, GA-17328, and GA-33112X, and by the Office of Naval Research under contract N00014-69-A-0200-1015.

## References

Axford, W. I., Magnetospheric convection, *Rev. Geophys. Space Phys.* **7**, 379–413 (1969).

Booker, H. G. and Gordon, W. E., The role of stratospheric scattering in radio communication, *Proc. IRE* **45**, 1223–1227 (1957).

Dolezalek, H., Discussion of the fundamental problem of atmospheric electricity, *Pageoph* **100**, 8–43 (1971).

Dungey, J. W., Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.* **6**, 47–48 (1961).

Gonzalez, W. D. and Mozer, F. S., A quantitative model for the potential resulting from reconnection with an arbitrary interplanetary magnetic field, Space Sciences Laboratory preprint, 44 pp. (1973).

Mozer, F. S., Balloon measurements of vertical and horizontal atmospheric electric fields, *Pageoph* **84**, 32–45 (1971).

Mozer, F. S., Analyses of techniques for measuring DC and AC electric fields in the magnetosphere, *Space Sci. Rev.* **14**, 272–313 (1973).

Mozer, F. S. and Serlin, R., Magnetospheric electric field measurements with balloons, *J. Geophys. Res.* **74**, 4739–4754 (1969).

Mozer, F. S. and Manka, R. H., Magnetospheric electric field properties deduced from simultaneous balloon flights, *J. Geophys. Res.* **76**, 1697–1712 (1971).

Park, C. G. and Dejnakarintra, M., The penetration of thundercloud electric fields into the ionosphere and magnetosphere; 1. Middle and subauroral latitudes, preprint from the Radioscience Laboratory, Stanford University, 26 pp. (1972).

Sagalyn, R. and Fitzgerald, D. R., Atmospheric electricity, in *Handbook of Geophysics and Space Environments*, Office of Aerospace Research, U.S. Air Force, 36 pp. (1965).

Woodman, R. F. and Guillen, A., Radar observations of wind and turbulence in the stratosphere and mesosphere, preprint of the Jicamarca Radio Observatory, 24 pp. (1972).

# The Average Auroral Zone Electric Field

F. S. Mozer and P. Lucht

Journal of Geophysical Research, vol. 79, no. 7, March 1, 1974

# The Average Auroral Zone Electric Field

F. S. MOZER AND P. LUCHT

*Physics Department and Space Sciences Laboratory*
*University of California, Berkeley, California 94720*

Four hundred and seventy-eight hours of electric field data obtained by the flight of 32 balloons between $L = 5.4$ and $L = 8.2$ have been averaged to determine mean properties of the auroral zone electric field. When it is mapped into the equatorial plane, the average $\mathbf{E} \times \mathbf{B}$ flow has a sunward component at all local times and produces a convection pattern consistent with that expected for flow around a rotating obstacle. The average auroral zone electric field strength increases by a factor of about 2 as $Kp$ increases from 0 to 6, but the larger-scale convection pattern is not greatly affected by the level of magnetic activity. The average electric field variation with $Kp$ is small compared to that caused by local turbulent fluctuations, and so local field strengths can be large during quiet periods or small during magnetic activity. The average electric field magnitude is enhanced by a factor of about 2 when the interplanetary magnetic field has a southward component; this finding provides further evidence in favor of the generation of the auroral zone electric field by magnetic field reconnection. The $y$ component of the interplanetary magnetic field causes the northern auroral zone electric field strength to be larger than average near local dawn, when $B_y > 0$, and larger than average near local dusk, when $B_y < 0$. These behaviors are understood in terms of a model in which the two-cell convection pattern is rotated about the sun-earth line by an amount that depends on the sign and magnitude of $B_y$. This model and the data suggest either that important parallel potential drops exist along auroral zone magnetic field lines or that simple conjugacy between the two hemispheres does not exist along such field lines.

Auroral ionospheric and magnetospheric electric fields have been measured by a variety of techniques (see *Mozer* [1973] for a review of the various types of measurements), and much information on certain aspects of their morphology has been obtained [*Mozer*, 1971*b*; *Cauffman and Gurnett*, 1972; *Doupnik et al.*, 1972; *Haerendel*, 1972; *McIlwain*, 1972]. However, relatively few attempts have been made to determine the diurnal, latitude, magnetic activity, and other dependencies of the average auroral zone electric field, probably because sufficiently large bodies of directly interpretable electric field data have not yet been obtained. Since one of the larger bodies of such data has resulted from the flight of 32 balloons between $L = 5.4$ and $L = 8.2$ [*Mozer and Serlin*, 1969; *Mozer and Manka*, 1971], the 478 hours of data thereby collected have been analyzed to determine the dependence of the average auroral zone electric field on these parameters. Whereas this body of data has been sufficient to elucidate certain general relationships that will be discussed below, it is important to note that further detailed analyses will probably require thousands of hours of data in order to average better over local turbulent fluctuations of these fields.

## DATA

Table 1 summarizes the locations and quantities of data collected from each of the seven sites in which balloon measurements of auroral zone ionospheric electric fields have been made. One-hour averages of the data from each of these flights are presented in a nonrotating frame of reference in Figure 1. Because of the scientific interest in nighttime measurements, data collection has been greatest at night. The largest sample of data obtained by any local time is the 31 hours acquired at 0200 LT, and the smallest amount is the 6 hours of data obtained at 1900 LT.

Figure 1 shows both a spread in the data at any local time and a general trend in the average field as a function of local time. Whereas the data are characterized as much by their spread as by their mean diurnal pattern, this turbulence will generally be eliminated from further discussion in the present paper by averaging over the data collected at any given local time. Properties of this turbulent behavior have been described elsewhere [*Mozer*, 1971*a*].

Copyright © 1974 by the American Geophysical Union.

## AVERAGE LARGE-SCALE MAGNETOSPHERIC ELECTRIC FIELD

In Figure 2 the hourly averages of the data of Figure 1 are plotted with error bars that represent the standard deviations of the means. The solid curves of this figure are smooth fits to the experimental data drawn with the constraint that the 24-hour average of the westward field component be zero in order to satisfy $\nabla \times \mathbf{E} = 0$. In agreement with expectation the electric field patterns of Figure 2 imply a sunward $\mathbf{E} \times \mathbf{B}$ convection, since the dominant field component is southward near dawn and northward near dusk. These data are also in excellent agreement with average flows deduced by barium cloud measurements [*Haerendel*, 1972], ground magnetometer observations [*Akasofu*, 1965], and radar incoherent backscatter [*Doupnik et al.*, 1972]. It is further noted that the southward field component is typically a factor of about 3 larger than the westward component, in agreement with all of the above techniques.

To deduce large-scale magnetospheric convection patterns from the average ionospheric data of Figure 2, it is convenient to map the ionospheric electric fields to the magnetic equatorial plane. On the assumption that the variation of the potential drop along neighboring magnetic field lines is small compared with the difference in potential between the lines, the ratios of the ionospheric to the equatorial electric field components are given by geometric factors that depend on the magnetic field configuration [*Mozer*, 1970] and are plotted in Figure 3 for an empirical magnetic field model [*Fairfield*, 1968] having magnetic field lines that enter the ionosphere at three latitudes. Because the equatorial

TABLE 1. Summary of Data

| Site | $L$ Value | Number of Flights | Hours of Data |
|---|---|---|---|
| Churchill | 8.2 | 12 | 160 |
| Yellowknife | 8.2 | 3 | 62 |
| Uranium City | 7.1 | 4 | 63 |
| Great Whale | 6.6 | 1 | 12 |
| Thompson | 6.4 | 8 | 126 |
| College | 5.4 | 3 | 41 |
| McMurray | 5.4 | 1 | 14 |

magnetic field strength varies rapidly with radius near local midnight in the field model (Figure 3, top), the ratio of the ionospheric to the equatorial electric field components, which are given in the bottom plots of Figure 3, depend critically on the latitude of the ionospheric end of the magnetic field line. Thus if the equatorial electric field strength is roughly independent of position near midnight over a region of a few earth radii near 10 $R_E$, the ionospheric field should vary rapidly with latitude, especially near midnight.

To determine whether such a rapid variation of ionospheric electric field strength with latitude is present in the experimental data, the curves of Figure 1 have been averaged over two latitude intervals and plotted in Figure 4. The higher-latitude data of this figure come from the first four sites listed in Table 1, and data during periods of extreme magnetic activity have been deleted from these averages. The differences between the higher- and the lower-latitude data are generally small near midnight and are not statistically significant at other times (near local noon there are only two lower-latitude measurements). Thus either the magnetic field model overestimates the radial gradient of the equatorial magnetic field at local midnight near 10 $R_E$, or there is a rapid radial increase of equatorial electric field strength that nearly compensates for the large magnetic field gradient. There is independent evidence that the former interpretation is more likely to be correct [*Sugiura*, 1972].

For the above reasons the 67° curves of Figure 3 have been used to map the mean electric fields of Figure 2 into the equatorial plane. The fields obtained by this mapping and given by the vectors in Figure 5 might be in error by factors of the order of 2 near midnight, owing to uncertainties in the magnetic field geometry. Nevertheless, certain general features of the deduced equatorial electric field pattern must still be valid. The field direction is generally from dawn to dusk, corresponding to sunward convection, and the field magnitude is typically about 1 mV/m. The field is small near 2200 and 1400 LT, corresponding to times when the average ionospheric field is nearly zero.

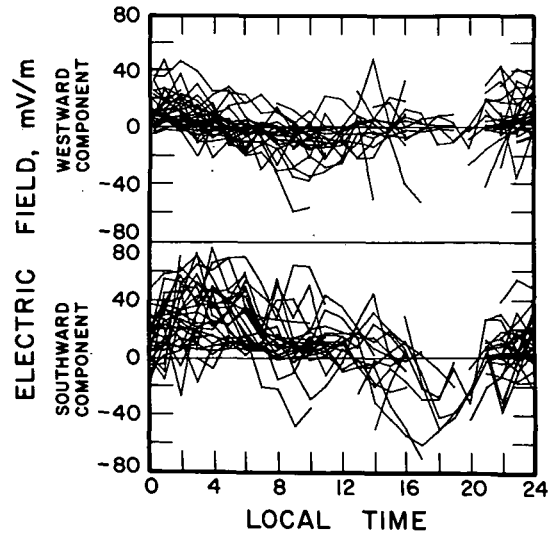



Fig. 1. One-hour averages of electric field components measured in a nonrotating frame of reference on 32 balloons flown in the auroral zone.

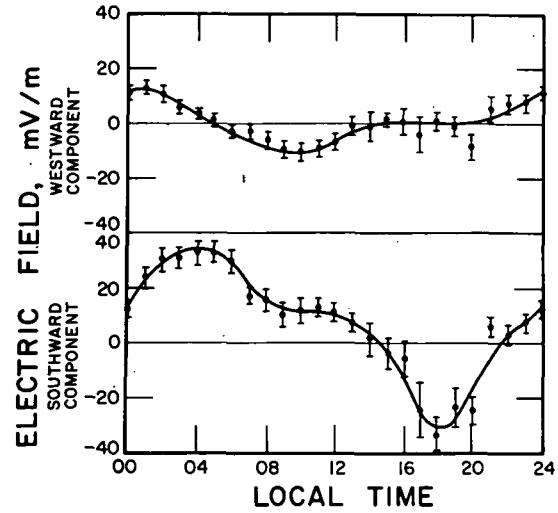



Fig. 2. Hourly averages of the data presented in Figure 1. The error bars are standard deviations of the means, and the solid curves are empirical fits to the data satisfying the constraint imposed by $\nabla \times \mathbf{E} = 0$ that the 24-h average westward field be zero.

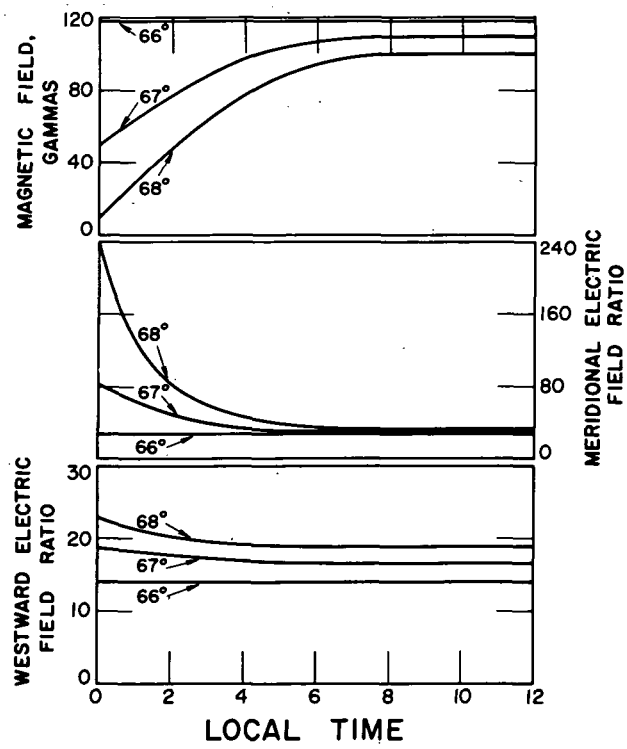



Fig. 3. (Top) Equatorial magnetic field strengths as a function of local time on field lines that intersect the ionosphere at three different latitudes. (Middle and Bottom) The ratios of ionospheric to equatorial electric field components on the same three magnetic field lines.

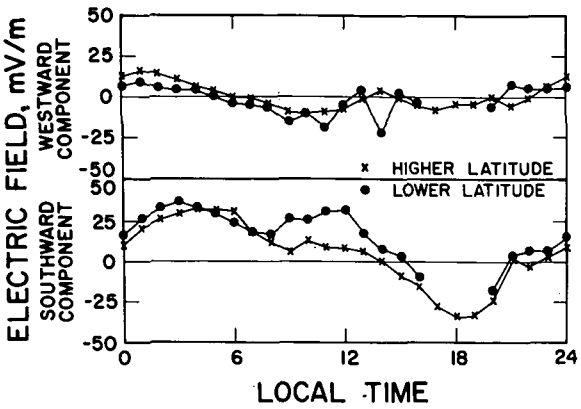



Fig. 4. Hourly averages of the electric field components measured at the higher- and lower-latitude sites.

Equipotentials of the electric field pattern are presented in Figure 5 as deduced from the measured vectors and the assumptions that the electric field is uniform across the tail at a tailward distance of 15 $R_E$, that the plasmapause is an electric equipotential, and that the magnetopause is not. Since the experimental data alone do not provide enough information for drawing equipotential patterns everywhere, assumptions other than these will produce different patterns. However, the experimental data do place important constraints on such patterns. The average total potential difference inside 8 $R_E$ cannot greatly exceed 30 kV. There are nulls in the flow near 1400 and 2200 LT.

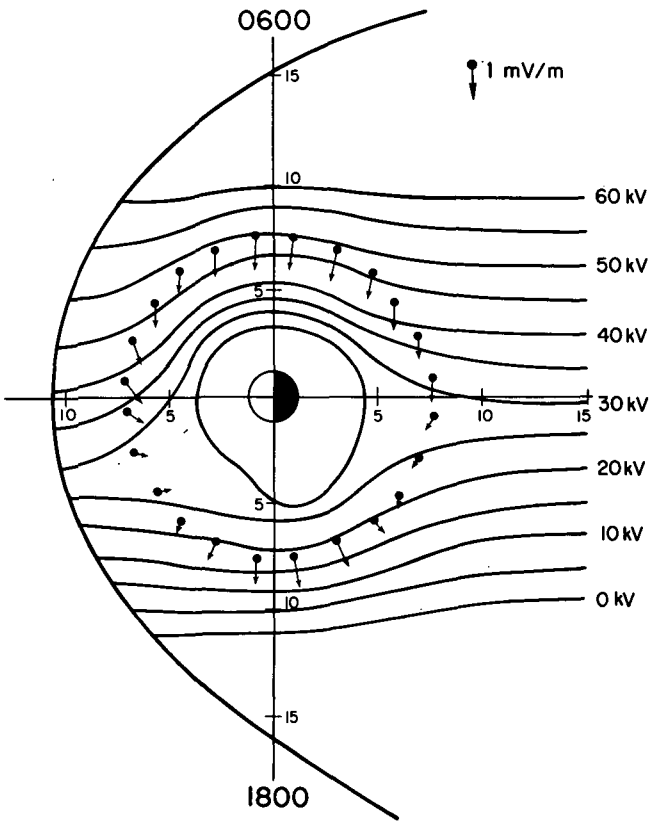



Fig. 5. Hourly averaged electric field vectors plotted in the equatorial plane of a nonrotating frame of reference, as viewed from above the north pole. The solid curves are electric equipotentials deduced from the experimental data on the assumption that the electric field is uniform across the tail at a tailward distance of 15 $R_E$ from the earth.

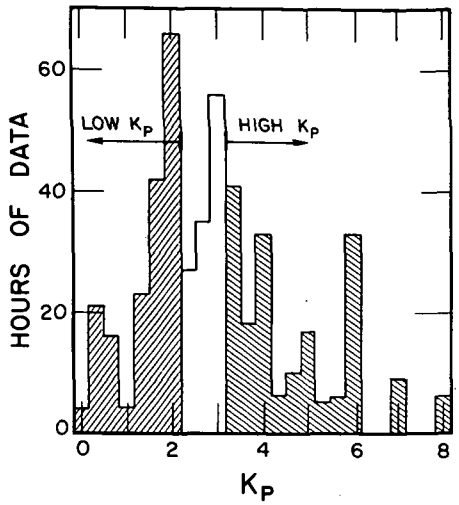



Fig. 6. Histogram of the amount of electric field data obtained as a function of *Kp*, shaded areas defining regions called low and high *Kp*.

The fact that the fields near dawn and dusk are larger than average limits the potential drop across an assumed non-equipotential plasmapause to about half of that existing if the electric field were uniform everywhere.

The experimental data and the above assumptions combine to suggest an equatorial potential drop of about 60 kV and an **E**×**B** potential flow in agreement with earlier theories (see review by *Axford* [1969]) and with expectations based on flow around a rotating impervious object. The equipotential plots of Figure 5 also generally agree with average equatorial field patterns deduced from plasma measurements [*McIlwain*, 1972, and revision, private communication, 1973].

## Variation of the Average Electric Field With Magnetic Activity

Figure 6 presents a histogram of the number of hours of electric field data collected as a function of the planetary *K* index *Kp*. The data have been divided into high- and low-*Kp* intervals as illustrated in Figure 6 such that roughly 40% of the data lies in each interval. The hourly averages of the electric field components for these high- and low-*Kp* intervals are plotted in Figure 7. Within the statistical accuracy of this figure the electric field is roughly twice as large during periods of enhanced *Kp* as it is when *Kp* is small, but the large-scale convection electric field pattern is not strongly dependent on the level of magnetic activity.

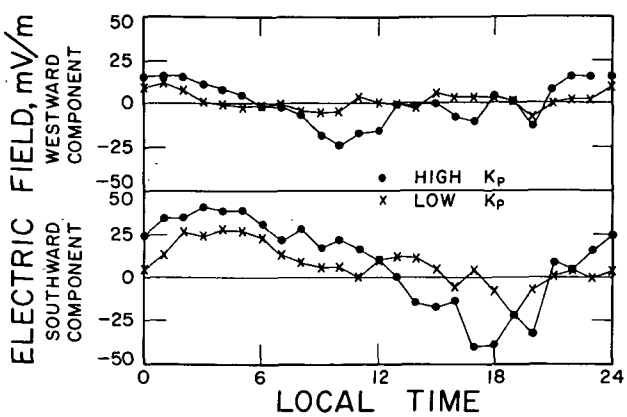



Fig. 7. Hourly averages of electric field components measured during periods of high and low *Kp*, as defined in Figure 6.

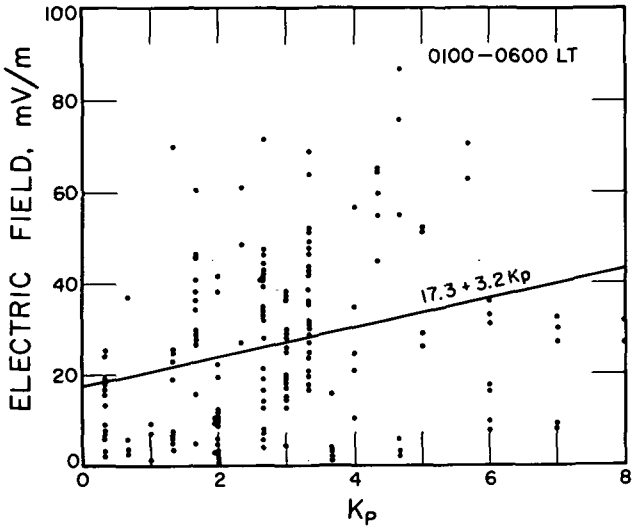



Fig. 8. Scatter diagram of 1-h averages of all electric field magnitudes measured between 0100 and 0600 LT, as a function of *Kp*. The solid curve is a linear least squares fit to the data.

The dependence of electric field strength on magnetic activity is further elucidated by the scatter plot of 1-h averages of the electric field strength versus *Kp* presented in Figure 8. Data from the local time interval of 0100–0600 have been selected for this study both because a large amount of data exists and because the average field strength varies by less than ±10% with local time over this interval. Although there is a small dependence of field strength on *Kp*, as illustrated by the linear least squares fit to the data, the major variations of the field are not explained in this way. In conclusion the average electric field strength varies by a factor of roughly 2 with magnetic activity, but this variation is less than that caused by local turbulent variations.

## DEPENDENCE OF THE AVERAGE ELECTRIC FIELD ON THE INTERPLANETARY MAGNETIC FIELD

One-hour averages of the electric field data have been divided into two subsets depending on the sign of the northward component of the interplanetary magnetic field, $B_z$, and the data in each subset have been averaged to produce the diurnal plots of Figures 9 and 10. The time delays Δt of these figures are the intervals between measurement of the interplanetary magnetic field and the ionospheric electric field. If attention is focused on the local time interval of 0000–1000 because differences between the two curves are statistically most significant during these times, then for any time delay between −4 and +4 hours it is observed that the average auroral zone electric field is roughly twice as large when the interplanetary magnetic field has a southward component as it is when the magnetic field is northward. This result is consistent with earlier electric field measurements [*Mozer*, 1971*b*; *Haerendel*, 1972] and offers strong evidence that the convection electric field is

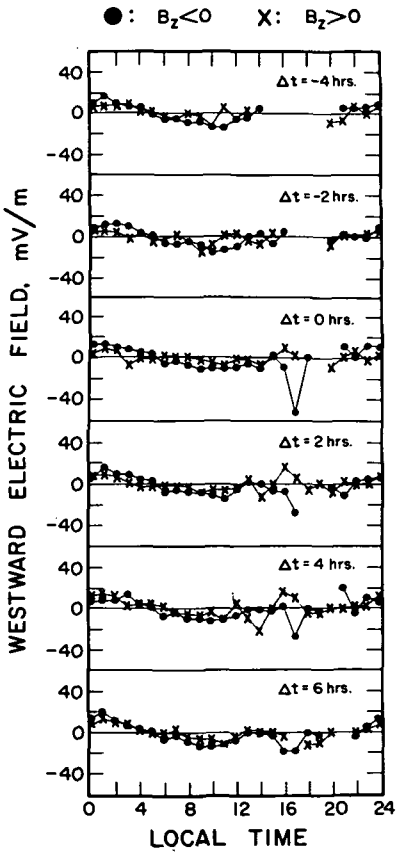



Fig. 9. Hourly averages of the westward electric field measured when the northward component of the interplanetary magnetic field was positive and when it was negative. The quantity Δt refers to the time delay between the measurement of the interplanetary magnetic field and that of the ionospheric electric field.

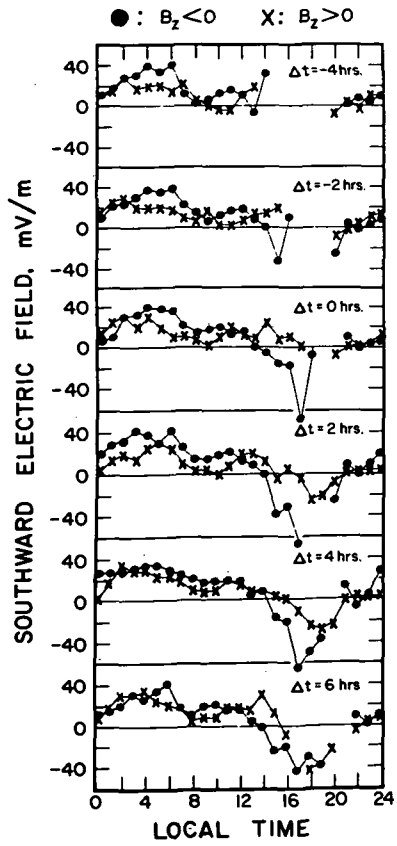



Fig. 10. Hourly averages of the southward electric field measured when the northward component of the interplanetary magnetic field was positive and when it was negative. The quantity Δt refers to the time delay between the measurement of the interplanetary magnetic field and that of the ionospheric electric field.

produced by magnetic field reconnection [*Dungey*, 1961]. It is also noted that the shape of the convection pattern is approximately independent of $B_z$ and is given by the curves of Figure 5 even though the average magnitude of the convection is related to the sign of $B_z$. That the shapes of the curves of Figures 9 and 10 are not strongly dependent on the assumed time delay is probably accounted for by the statistical uncertainties in the curves and by the observation that the correlation time scale of the interplanetary magnetic field is apparently several hours.

The dependence of the average auroral zone electric field on the $y$ component of the interplanetary magnetic field is given in Figures 11 and 12. For any time delay less than about 8 hours the southward component of electric field is largest in the postmidnight time interval when $B_y > 0$. The curves of Figure 12 also suggest with less statistical significance that the magnitude of the southward field component is largest near dusk when $B_y < 0$. These results may be understood in terms of a reconnection model (W. Gonzalez and F. S. Mozer, unpublished manuscript, 1973) in which the line of reconnection depends on the direction of the interplanetary magnetic field. For a purely southward magnetic field the reconnection line is in the equatorial

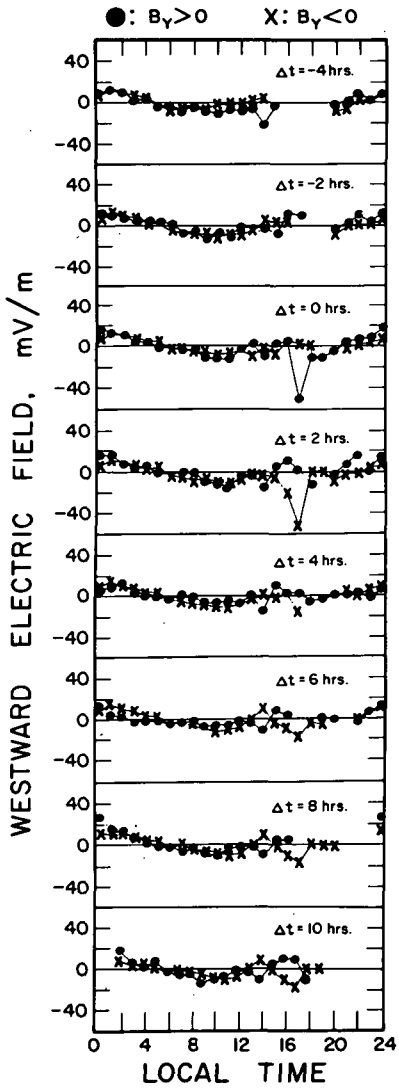



Fig. 11. Hourly averages of the westward electric field measured when the $y$ component of the interplanetary magnetic field was either positive or negative.

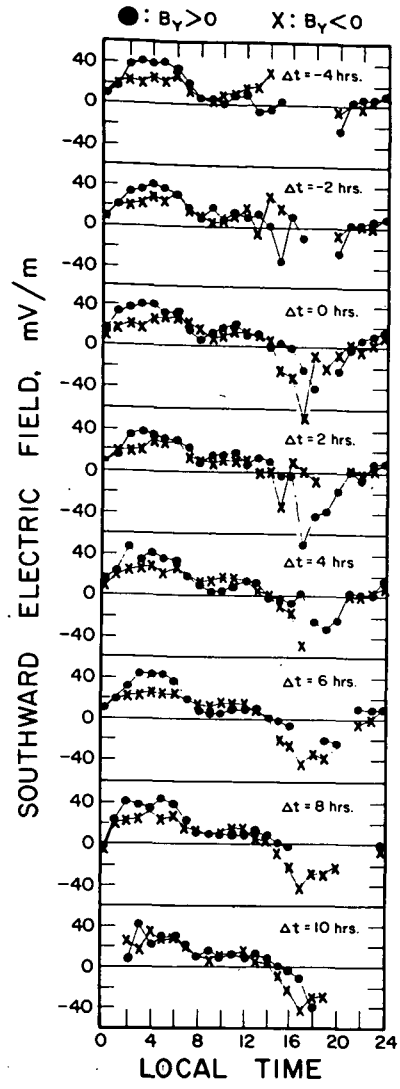



Fig. 12. Hourly averages of the southward electric field measured when the $y$ component of the interplanetary magnetic field was either positive or negative.

plane, and the reconnection line rotates out of this plane as the interplanetary field rotates about the sun-earth line. Since the postreconnection flow is initially perpendicular to the reconnection line, the two-cell magnetospheric convection pattern tends to rotate about the sun-earth line as the interplanetary magnetic field direction rotates. This rotation causes a larger return flow at a fixed northern hemisphere sub-polar cap latitude at dawn, when $B_y > 0$, and at dusk, when $B_y < 0$. This rotation of the two-cell convection pattern has been observed in polar cap electric field data [*Heppner*, 1972; *Mozer et al.*, 1974] as well as in the southward electric field component of Figure 12. It is also consistent with ground observations of magnetic field variations in the polar cap [*Svalgaard*, 1968; *Mansurov*, 1969; *Berthelier and Guerin*, 1972]. That the sign $B_y$ has little effect on the westward field component of Figure 11 is understood from the facts that the major effect is expected near local dawn and dusk and that these are times when the average westward electric field is zero.

Since the hypothesized rotation of the two-cell convection pattern occurring when $B_y > 0$ causes an enhanced electric field near dawn at a fixed latitude in the northern hemisphere and a reduced field at the same time and latitude in

the southern hemisphere, the ionospheric electric fields in the dawn-dusk magnetic meridional plane at equal northern and southern hemisphere auroral zone latitudes are not equal. Figure 12 suggests that the magnitude of this difference is a reasonable fraction of the larger field strength. This result may be due to a lack of conjugacy at auroral zone latitudes near dawn and dusk or may be due to parallel potential drops that are comparable to the ~60-kV average potential drop across the magnetosphere. There is considerable experimental [*Mozer and Fahleson*, 1970; *Kelley et al.*, 1971*a*, *b*; *Mozer et al.*, 1973] and theoretical [*Kindel and Kennel*, 1971; *Coroniti and Kennel*, 1972] evidence that parallel potential drops of tens of kilovolts exist in the magnetosphere. Such results are consistent with many laboratory experiments [*Hamberger and Jancarik*, 1970] and plasma theoretical calculations [*Buneman*, 1958].

*Acknowledgments.* This work was supported by the Atmospheric Sciences Section of the National Science Foundation under grants GA 1317, GA 11259, GA 28307, GA 17328, and GA 33112X, by the Office of Naval Research under contracts N00014-67-A0112-0068 and N00014-69-A-02000-01015, and by NASA under grant NAS 9-9502.

* * *

The Editor thanks J. R. Doupnik and A. Pedersen for their assistance in evaluating this paper.

## REFERENCES

Akasofu, S.-I., Dynamic morphology of auroras, *Space Sci. Rev.*, *4*, 498, 1965.

Axford, W. I., Magnetospheric convection, *Rev. Geophys. Space Phys.*, *7*, 379, 1969.

Berthelier, A., and C. Guerin, Influence of the polarity of the interplanetary magnetic activity at high latitudes, *Rep. GRI/NT/87*, 11 pp., Groupe de Rech. Ionos., Centre Nat. de la Rech. Sci., Saint-Maur, France, 1972.

Buneman, O., Instability, turbulence, and conductivity in current-carrying plasma, *Phys. Rev. Lett.*, *1*, 8, 1958.

Cauffman, D. P., and D. A. Gurnett, Satellite measurements of high latitude convection electric fields, *Space Sci. Rev.*, *13*, 369, 1972.

Coroniti, F. V., and C. F. Kennel, Polarization of the auroral electrojet, *J. Geophys. Res.*, *77*, 2835, 1972.

Doupnik, J. R., P. M. Banks, M. J. Baron, C. L. Reno, and J. Petriceks, Direct measurement of plasma drift velocities at high magnetic latitudes, *J. Geophys. Res.*, *77*, 4268, 1972.

Dungey, J. W., Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.*, *6*, 47, 1961.

Fairfield, D. H., Average magnetic field configuration of the outer magnetosphere, *J. Geophys. Res.*, *73*, 7329, 1968.

Haerendel, G., Plasma drifts in the auroral ionosphere derived from barium releases, in *Earth's Magnetospheric Processes*, edited by B. M. McCormac, p. 246, D. Reidel, Dordrecht, Netherlands, 1972.

Hamberger, S. M., and J. Jancarik, Dependence of anomalous conductivity of plasma on the turbulent spectrum, *Phys. Rev. Lett.*, *25*, 999, 1970.

Heppner, J. P., Polar cap electric field distributions related to the interplanetary magnetic field direction, *J. Geophys. Res.*, *77*, 4877, 1972.

Kelley, M. C., F. S. Mozer, and U. V. Fahleson, Electric fields in the nighttime and daytime auroral zone, *J. Geophys. Res.*, *76*, 6056, 1971*a*.

Kelley, M. C., F. S. Mozer, G. Haerendel, U. V. Fahleson, and A. Kavadas, Rocket dc electric field measurements in a 1000-gamma magnetic substorm, *Eos Trans. AGU*, *52*, 329, 1971*b*.

Kindel, J. M., and C. F. Kennel, Topside current instabilities, *J. Geophys. Res.*, *76*, 3055, 1971.

Mansurov, S. M., New evidence of a relationship between magnetic fields in space and on earth, *Geomagn. Aeron.*, *9*, 622, 1969.

McIlwain, C. E., Plasma convection in the vicinity of the geosynchronous orbit, in *Earth's Magnetospheric Processes*, edited by B. M. McCormac, p. 268, D. Reidel, Dordrecht, Netherlands, 1972.

Mozer, F. S., Electric field mapping in the ionosphere at the equatorial plane, *Planet. Space Sci.*, *18*, 259, 1970.

Mozer, F. S., Power spectra of the magnetospheric electric field, *J. Geophys. Res.*, *76*, 3651, 1971*a*.

Mozer, F. S., Origin and effect of electric fields during isolated magnetospheric substorms, *J. Geophys. Res.*, *76*, 7595, 1971*b*.

Mozer, F. S., Analysis of techniques for measuring dc and ac electric fields in the magnetosphere, *Space Sci. Rev.*, *14*, 272, 1973.

Mozer, F. S., and U. V. Fahleson, Parallel and perpendicular electric fields in an aurora, *Planet. Space Sci.*, *18*, 1563, 1970.

Mozer, F. S., and R. H. Manka, Magnetospheric electric field properties deduced from simultaneous balloon flights, *J. Geophys. Res.*, *76*, 1697, 1971.

Mozer, F. S., and R. Serlin, Magnetospheric electric field measurements with balloons, *J. Geophys. Res.*, *74*, 4739, 1969.

Mozer, F. S., F. H. Bogott, and B. Tsurutani, Relations between ionospheric electric fields and energetic trapped and precipitating electrons, *J. Geophys. Res.*, *78*, 630, 1973.

Mozer, F. S., W. D. Gonzalez, F. H. Bogott, M. C. Kelley, and S. Schutz, High-latitude electric fields and the three-dimensional interaction between the interplanetary and the terrestrial magnetic fields, *J. Geophys. Res.*, *79*, 56, 1974.

Sugiura, M., The ring current and associated phenomena, in *Critical Problems of Magnetospheric Physics*, edited by E. R. Dyer, p. 195, National Academy of Sciences, Washington, D.C., 1972.

Svalgaard, L., Sector structure of the interplanetary magnetic field and daily variations of the geomagnetic field at high latitudes, *Geophys. Pap. R-6*, Dan. Meteorol. Inst., Copenhagen, Denmark, 1968.

(Received May 29, 1973;
accepted November 7, 1973.)

A74-16900

# Gas-Rich Meteorites: Possible Evidence for Origin on a Regolith

Abstract. *The asymmetry of irradiation features of grains in the Kapoeta and Fayetteville meteorites suggests irradiation on a regolith before meteorite formation. Chondrules and broken grains require approximately $10^4$ years of irradiation time between formation or fracturing and compaction into the meteorite. Shock erasure of tracks from irradiated Kapoeta feldspars requires a severe shock event during or after meteorite formation.*

Gas-rich meteorites contain large amounts of rare gases (*1*) which have abundance ratios similar to those in the sun (*2*). Controlled etching experiments have shown that the "solar" gas resides in the outer few micrometers of individual grains and have led to the suggestion that the gases resulted from a low-energy particle irradiation (*3*) similar to an irradiation by present-day solar flare nuclei. This theory was corroborated by the discovery that a fraction of the grains in several gas-rich meteorites have high densities of particle tracks and steep track density gradients (*4*). Such track features could only have been produced by a low-energy particle irradiation of these grains as individuals with negligible shielding and could not result from cosmic-ray nuclei stopping in them at their present sites within the meteorite. Crystals that exhibit high track densities have been termed "track-rich" grains.

Irradiation geometry features of these grains provide evidence about the environment at the time of their irradiation. In particular, the discoverers of the track-rich grains reported that the irradiation features are apparently symmetrical (that is, the irradiation dose is the same at all faces of a given grain) and inferred that the grains were probably irradiated in free space (*4*). We report here several new features of track-rich grains observed during a detailed investigation of two well-known gas-rich meteorites, Kapoeta and Fayetteville. High-resolution electron microscope methods, described elsewhere (*5–7*), have increased our capability to resolve tracks at densities an order of magnitude or more higher than was previously possible. Here we outline briefly the new observations and their implications; a more detailed treatment has been presented elsewhere (*7*).

Figures 1, 2, and 3 show examples of the features described in this report. Figure 1 is a scanning electron microscope photograph of a track-rich chondrule-like olivine grain in situ in a section of Fayetteville chemically etched to reveal tracks in olivine. Although we have observed several intact irradiated chondrules in this meteorite, the importance of this particular observation is that this object has been broken and a track density gradient exists at the broken edge (inset), an indication that the low-energy particle irradiation took place after fragmentation. If the flux of low-energy particles which irradiated the track-rich chondrules in Fayetteville was similar to the present-day flux from solar flares, the observed track densities require $\sim 10^4$ years of irradiation time. Similar times were suggested previously (*4*) for track-rich grains from several gas-rich meteorites. The similarity of solar conditions at the time of irradiation and the present cannot be checked directly. Under the assumption, however, $10^4$ years is a lower limit on the time between chondrule formation and meteorite compaction. Furthermore, the track density gradients are so steep (*7*) that the thickness of gas or solid matter between the sun and the irradiated chondrules (and grains) could not have exceeded $\sim 10^{-3}$ g/cm$^2$.

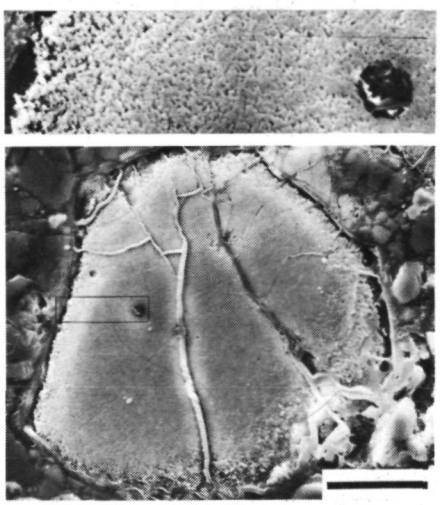


Fig. 1. Scanning electron microscope photograph of a track-rich olivine grain in a Fayetteville section. This broken olivine chondrule was irradiated after fracturing as shown by the track gradient at the broken edge. The area in the box is shown above the main illustration at × 5.1 greater magnification. The track density at the broken edge is $\sim 10^{10}$ tracks per square centimeter. The scale bar is 20 μm for the main photograph.

Figure 2 shows a track-rich feldspar grain in situ in a section of Kapoeta. The grain exhibits a feature which we have observed frequently in both Fayetteville and Kapoeta, namely, nonuniform track density around the grain borders. The variation is regular, with the highest density along one edge decreasing smoothly toward an opposite edge. The highest and lowest edge densities differ by approximately a factor of 3. Most track-rich grains in Fayetteville exhibit a similar nonuniformity (*7*).

The track densities at grain edges ($\sim 10^9$ tracks per square centimeter in Kapoeta and $\sim 10^{10}$ in Fayetteville) are much too high to resolve by optical microscopy and require very light etching combined with high-resolution techniques for their measurement. Thus it is not surprising that the nonuniformity was overlooked in the original descriptions of track-rich grains (*4*) which were, for the most part, based on optical microscopy. About 90 percent of the grains we observed have higher track densities at all edges than at their centers. This is in contrast to the case of lunar soil grains of similar size, which commonly have high track densities on only one side (*8, 9*).

The nonuniform irradiation features provide an important point for speculation about the environment of gas-rich meteorite formation, although the implications are not immediately obvious. It seems clear that such features could not be a *primary* result of "free space" irradiation of individual grains. Postirradiation erosion, or attrition during meteorite compaction, could result in nonuniform irradiation features, although it is difficult to see how this could explain the observed regular change in track density around the grain periphery. Arrhenius (*10*) has suggested irradiation in space in a loose agglomeration of small grains. The features in Kapoeta and Fayetteville are reminiscent of irradiated lunar fines, although, as pointed out above, the lunar grains commonly have high track densities on only one side. This suggests a higher "gardening" rate and possibly a lower gravitational field in the case of the meteorites, which would be compatible with earlier suggestions that the surfaces of small bodies are the sites of gas-rich meteorite formation [for example, see (*11*)].

Figure 3 is a micrograph of a shad-

Reprinted from SCIENCE, 11 January 1974, volume 183, pages 73-74

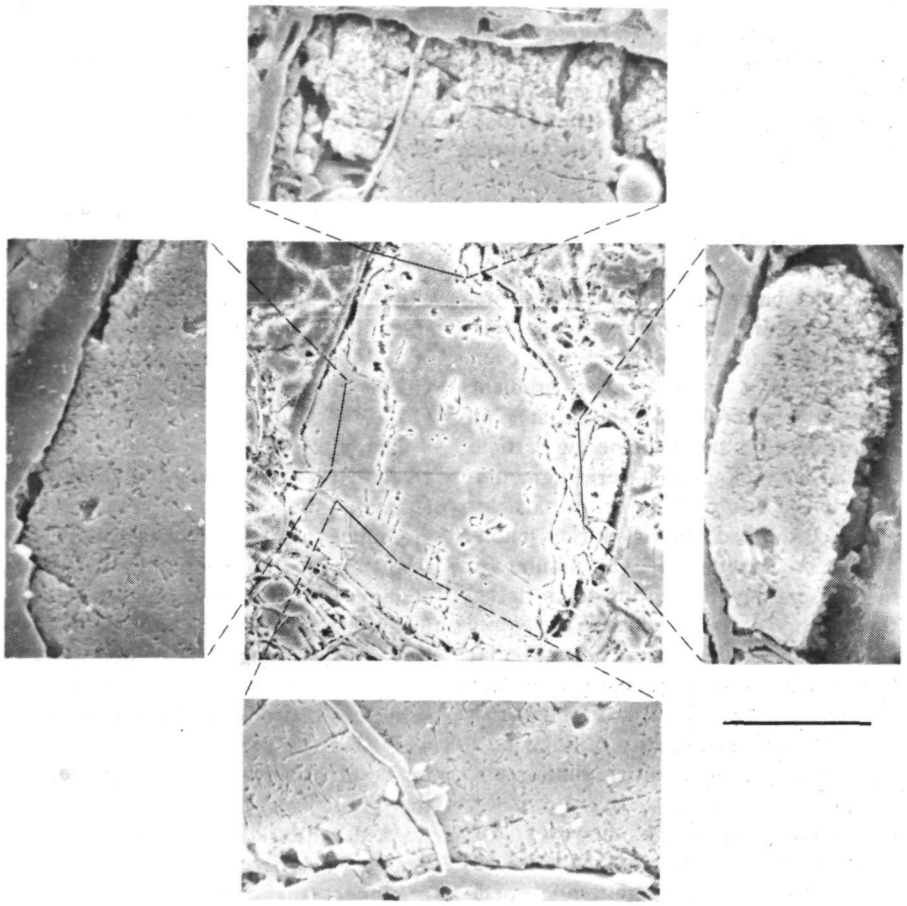


Fig. 2. Track-rich feldspar grain in a Kapoeta section, seen with the scanning electron microscope. Edges are shown at × 3.8 the magnification of the central grain to show the anisotropic irradiation effects. The scale bar is 60 μm for the central photograph. The track density is highest at the top and right side of the grain, lowest on the left side.

owed carbon replica (*5*) of an etched feldspar grain in a section of Kapoeta. Near the center of this elongate grain are areas where the tracks are short and the track density is lower than in adjacent regions. These areas are also characterized by narrowly spaced parallel markings which appear to be slip planes of the type described by Fleischer *et al.* (*12*). Thus this grain preserves a record of a shock event which occurred after its irradiation, since all tracks in the deformed areas are shortened. Evidence for extensive shock has been observed in Kapoeta by others (*13*); however, we had previously assumed that particle tracks would not survive the shocking and used the presence of track-rich grains as evidence for an absence of shock events during or after the compaction of Kapoeta (*8*). The results presented here indicate that a shock event affected Kapoeta components after irradiation of the track-rich grains, possibly during compaction of the meteorite. The event partially erased tracks in some grains and may have entirely removed them from others. This may partly explain the observed low abundance of track-rich grains in Kapoeta (~ 8 percent) as compared with Fayetteville (~ 30 percent).

Particle track features reported here, in addition to the steepness of the observed track density gradients in Fayetteville grains (*7*), can be used to place some limits on conditions during the history of Kapoeta and Fayetteville.

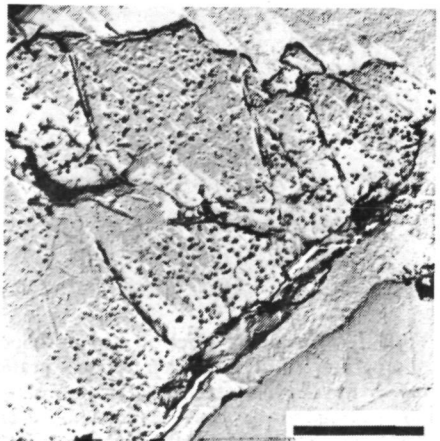


Fig. 3. Transmission electron microscope photograph of a shadowed carbon replica of an irradiated Kapoeta feldspar showing shock features as explained in the text. The scale bar is 4 μm. Tracks in this photograph appear as black, projecting objects with white shadows.

Irradiation on a small body would require a negligible atmosphere of $\lesssim 10^{-3}$ g/cm$^2$. If one postulates that these meteorites were formed in a jet stream of the type envisioned by Alfvén and Arrhenius (*14*), then gas pressure in the swarm must have been less than $\sim 10^{-9}$ atm. Either prior to or during the irradiation process, relative grain velocities must have been high enough to fragment chondrules. Although it has been previously mentioned that there are track-rich chrondrules in Weston (*4*), it has not been emphasized that this finding requires chondrule formation (and, as reported here, chondrule fracturing) $\geqslant 10^4$ years before meteorite compaction. Finally, nonuniform irradiation effects must be a result of any model for gas-rich meteorite formation.

D. MACDOUGALL
R. S. RAJAN
P. B. PRICE
*Department of Physics, University of California, Berkeley 94720*

### References and Notes

1. E. K. Gerling and L. K. Levskii, *Dokl. Akad. Nauk SSSR* **110**, 750 (1956).
2. P. Signer and H. E. Suess, in *Earth Sciences and Meteoritics*, J. Geiss and E. D. Goldberg, Eds. (North-Holland, Amsterdam, 1963), p. 241.
3. H. E. Suess, H. Wänke, F. Wlotzka, *Geochim. Cosmochim. Acta* **28**, 595 (1964); H. Wänke, *Z. Naturforsch. A* **20**, 946 (1965); P. Eberhardt, J. Geiss, N. Grögler, *Earth Planet. Sci. Lett.* **1**, 7 (1966).
4. D. Lal and R. S. Rajan, *Nature (Lond.)* **223**, 269 (1969); P. Pellas, G. Poupeau, J. C. Lorin, H. Reeves, J. Audouze, *ibid.*, p. 272.
5. D. Macdougall, D. Lal, L. Wilkening, S. Liang, G. Arrhenius, A. S. Tamhane, *Geochem. J.* **5**, 95 (1971).
6. D. Macdougall, R. S. Rajan, I. D. Hutcheon, P. B. Price, *Geochim. Cosmochim. Acta*, in press.
7. R. S. Rajan, *ibid.*, in press.
8. G. Arrhenius *et al.*, *ibid.* **3** (Suppl. 2), 2583 (1971).
9. D. Macdougall, thesis, University of California, San Diego (1972); ——— and R. S. Rajan, unpublished data.
10. G. Arrhenius, personal communication.
11. E. Anders, *Space Sci. Rev.* **3**, 583 (1964).
12. R. L. Fleischer, G. M. Comstock, H. R. Hart, Jr., *J. Geophys. Res.* **77**, 5050 (1972); R. L. Fleischer and H. R. Hart, Jr., *Gen. Elec. Tech. Inf. Ser. Rep. 72CRD300* (1972).
13. K. Fredriksson and K. Keil, *Geochim. Cosmochim. Acta* **27**, 717 (1963).
14. H. Alfvén and G. Arrhenius, *Astrophys. Space Sci.* **8**, 338 (1970); *ibid.* **9**, 3 (1970).
15. This work was supported by NASA grant NGL 05-003-410.

11 August 1973; revised 15 October 1973

Copyright © 1974 by the American Association for the Advancement of Science

D11
D12

IRRADIATION RECORDS IN ORANGE GLASS AND TWO BOULDERS FROM APOLLO 17.
D. Macdougall, I.D. Hutcheon and P.B. Price, Physics Dept., Univ. of Calif., Berkeley, Ca. 94720

Orange glass from subsurface samples 74220 and 74261 - The possibility that the bright orange material at station 4 represented recent lunar volcanism drew widespread interest to this sample. We have measured particle track densities, diameter distributions and annealing behavior as well as uranium concentration in the orange glass.

The major particle track features of the glass are: (1) In 74220 the total measured track density in individual spheres ranges between $1.5\times10^5$ and $10^8$ (median $5\times10^5$) for 16 spheres. For 15 spheres from 74261 the range is $2.2\times10^5$ to $1.3\times10^7$ (median $4.5\times10^5$). (2) Roughly 10% of the spheres in each sample have track density gradients indicative of unshielded exposure on the lunar surface. (3) For fixed etching conditions, the measured track diameter distributions in individual spheres vary depending upon track density. Spheres with low track densities ($\langle(1\text{-}3)\times10^5$ cm$^{-2}$) have the largest track diameters (see Fig. 1). (4) Annealing experiments show that iron nuclei tracks are partially erased by heating at 200°C for 1 hour, and that fission tracks (both spontaneous and induced) are completely erased by heating at 450°C for 1 hour. These conclusions were drawn from measured track density distributions for groups of spheres heated at various temperatures between 150 and 500°C for 1 hour. Relative to the distribution for unheated spheres, samples heated at 150 and 200° show a progressive shift toward lower track densities (fading of cosmic ray Fe tracks); between 200 and 300° the distribution is relatively stable (mostly unaffected fission tracks) while at 350° and above another progressive shift toward lower track densities is noted (fading of fission tracks). No natural tracks remain after 1 hour at 450°C; neutron-induced fission tracks also anneal under these conditions. (5) The uranium concentration in the glass, measured by counting induced fission tracks in an adjacent mica detector, is quite homogeneous with an area-weighted average (15 spheres) of 155 ppb for 74220 and 159 ppb (15 spheres) for 74261.

From the above the following conclusions can be drawn. The diameter distributions and annealing data indicate that most (50-100%) tracks in spheres with densities of $(1\text{-}3)\times10^5$ tracks cm$^{-2}$ are fission tracks. This is also borne out by the fact that neutron-induced fission tracks in the same spheres have a diameter distribution similar to (but on the average ~10% larger than) the natural tracks. This slight size difference may indicate minor annealing (up to ~20% based on typical calibration curves (1)) of the spontaneous fission tracks at lunar surface temperatures. The expected fission track density for 0.16 ppm uranium and an age of $3.7\times10^9$ y (2) is $2\times10^5$ tracks/cm$^2$, taking into account the experimentally determined etching efficiency of 0.6. A 20% annealing correction would reduce this to $1.6\times10^5$. Thus, contrary to some other interpretations (3) we believe the fission track record in the orange spheres to be compatible with the $^{40}$Ar-$^{39}$Ar age.

Soil 74220 was collected from a distinct layer extending from approximately 5 to 7.5 cm below the lunar surface. Soil 74261 was collected from the gray soil zone at the end of the same trench. Both samples may include

material from the overlying surface soil; in fact this is the probable source for the ~10% of solar flare irradiated orange glass in each sample. Since the soil is well stratified with distinct upper and lower boundaries, it has probably never been vigorously mixed. Using the median track density, the in-situ galactic cosmic ray exposure age is $2.5 \times 10^6$ yrs for soil 74220. (A correction of $2 \times 10^5/cm^2$ has been subtracted from the median density given above in view of the fission track contribution.) Because of the much longer (~30 My) Ne exposure age (4) the orange glass must have been within ~1 m of the lunar surface but below ~15 cm for ~30 My at some time in its history, and at its present position for only $2.5 \times 10^6$ yrs.

Apollo 17 boulders - Our interest in the Apollo 17 boulders has been directed toward an understanding of individual boulder history through determination of particle track exposure ages, erosion rates, etc., and also toward a search for pre-compaction irradiation effects and breccia formation effects in general through the use of high voltage electron microscopy. The data presented below were gleaned for the most part from small, sometimes undocumented chips in a preliminary allocation. It is expected that much more comprehensive information will be available at the time of the Fifth Lunar Science Conference.

One unexplained phenomenon which has hindered the track studies up to the present is the peculiar etching behavior of many of the feldspar grains, particularly in our samples of 72255. After standard etching procedures (using an etching solution of 1gNaOH:2g$H_2O$) feldspar surfaces are frequently irregular and 'bumpy', presumably due to non-uniform dissolution of the surface, making track observations difficult and perhaps introducing errors if track etching is as irregular as is surface dissolution. Optical microscope observations are possible in areas of low track density if surfaces are lightly polished to remove irregularities after etching; however, scanning electron microscopy of areas with densities $\gtrsim 10^6 cm^{-2}$ is very difficult. Although we have not noticed this effect in feldspars from other missions, it is apparently widespread among Apollo 17 samples as we have observed similar behavior in numerous surface soil samples, in other boulder samples, in basalt 71055 and occasionally in Apollo 17 deep core samples. The peculiar etching behavior may be related to small-scale chemical irregularities. Shock would appear to be ruled out as an explanation since shock effects are common in feldspars of similar bulk composition from other missions, yet the etching abnormality seems restricted to Apollo 17 samples.

Boulder 1, Station 2 - 72255: Preliminary studies of two interior pieces from this sample indicate an upper limit to the track exposure age of 15-20 My. Exposure ages determined by track methods, in the context of the large boulders, require some explanation. They do not necessarily bear any relation to the time at which the boulder rolled to its present location. Lunar surface photographs show considerable debris, including large fragments, surrounding the boulders. Because of the steep depth dependence of the particle track production rate, the surface under a spalled off fragment of typical thickness 10 cm is essentially virgin from a track point of view. Thus we expect that particle track exposure ages determined at different locations on the same large boulder will reflect the local rate of large-scale discontin-

March, 1974 (to be published in the Proc. Fifth Lunar Science Conf.)

# IRRADIATION RECORDS IN ORANGE GLASS

Macdougall, D. et al.

uous erosion, and will not necessarily be the same.

Boulder 2, Station 2 - 72315: Our measurements of a steep track density profile in a single feldspar crystal from this sample are documented in an accompanying abstract (5). An exceptionally young solar flare track exposure age of $10^5$ y has been calculated. 72335: We have examined only a 60 mg undocumented chip, supposedly a surface sample, although without a discernible track density gradient at the edges examined. However, a gradient typical of solar flare track gradients was observed to extend over several feldspar grains in the chip interior. Maximum observed track densities were $>5\times10^8$ cm$^{-2}$, falling to a minimum of $\sim3\times10^7$ and rising again. We have observed large variations ($\sim$a factor of 6) in track density between adjacent olivine grains and variations in track density of more than 20 x between adjacent olivine and feldspar grains. Possible explanations: (i) 72335 might contain a few pre-compaction irradiated grains. (ii) The interior track density gradient and the occasional high track density feldspar may be due to an unusual exposure geometry created by the presence of numerous nearby vugs in the exposed surface. (iii) Track density differences among adjacent olivines may be due to shock or thermal annealing. 72395: We have examined a documented interior slab that was cut roughly parallel to and ~3 cm below the exposed boulder surface. Track densities in feldspar range between $2\text{-}5\times10^6$ cm$^{-2}$. A tentative exposure age of 10-20 My for this part of the boulder can be assigned.

High voltage electron microscope studies of the boulder samples will proceed as soon as the refurbishment of the 650kV microscope is completed.

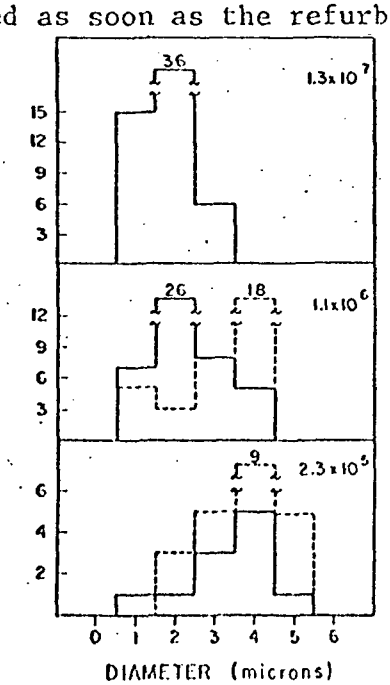



Fig. 1 - Track diameter histogram for spheres of varying track densities from 74220. Neutron-induced fission track diameters for the same spheres are shown as dotted lines.

References

1. D. Storzer and G.A. Wagner, Earth Planet. Sci. Lett. 5, 463 (1969).
2. L. Husain and O.A. Schaeffer, Science 180, 1358 (1973).
3. D. Storzer et al., Proc. Fourth Lunar Science Conference 3, 2363 (1973).
4. T. Kirsten et al., Trans. Am. Geophys. Union 54, 595 (1973).
5. I.D. Hutcheon, D. Macdougall, P.B. Price, F. Hörz, D. Morrison and E. Schneider, this volume.

THE ASTROPHYSICAL JOURNAL, 188:349–352, 1974 March 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

A74-32103

# ANALYTIC VARIATIONAL CALCULATION OF THE GROUND-STATE BINDING ENERGY OF HYDROGEN IN INTERMEDIATE AND INTENSE MAGNETIC FIELDS

LANCE W. WILSON
Department of Physics, University of California, Berkeley, California 94720
*Received 1973 April 23; revised 1973 June 25*

ABSTRACT

A "prescription" of simple functions is given for an analytic variational calculation of the binding energy of the hydrogen ground state. The calculation still yields a relatively smooth transition between intermediate and intense fields. An explicit calculation of the binding energy as $B \to \infty$ is given.

*Subject heading:* magnetic fields

## I. INTRODUCTION

The probable existence of intense magnetic fields ($\sim 10^{12}$ gauss) near pulsars and the discovery of intermediate ($\sim 10^{7}$ gauss) magnetic fields in white dwarfs (Kemp 1970; Kemp *et al.* 1970) have stimulated interest in the properties of matter in such high magnetic fields. In particular, the effect of a uniform magnetic field on the binding energy of the hydrogen ground state has been studied by several groups. Rajagopal *et al.* (1972) did variational calculations for intermediate fields ($10^{8}$ to $3 \times 10^{10}$ gauss); while Yafet, Keyes, and Adams (1956), Cohen, Lodenquai, and Ruderman (1970), and, more recently, Canuto and Kelly (1972), have done variational calculations emphasizing intense fields ($3 \times 10^{10}$ to $> 10^{12}$ gauss). Note that this work is concerned only with variational calculations that can be done analytically; however, if even higher accuracy for the transition region is desired, several groups have done rather extensive numerical calculations. For details, see Cabib, Fabri, Fiorio (1971, 1972), Pokatilov and Rusanov (1969), Larsen (1968), Baldereschi and Bassani (1970), and Callaway (1972). The first purpose of this work is to provide an analytically simple wave-function "prescription" which still gives a smooth transition between intermediate and intense fields.

Both Cohen *et al.* and Canuto and Kelly refer to analytical solutions for the one-dimensional hydrogen atom. However, neither work shows that the ground-state binding energy approaches the expected one-dimensional limit as $B \to \infty$. Thus, the second purpose of this work is to provide an explicit calculation of the ground-state binding energy as $B \to \infty$, at least for one type of trial function.

## II. REVIEW OF PREVIOUS VARIATIONAL CALCULATIONS

With the uniform field in the $z$-direction, the Pauli equation in spherical coordinates becomes

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + \omega_l L_z + \tfrac{1}{2}m\omega_l^2 r^2 \sin^2\theta + \hbar\omega_l\sigma_z - \frac{e^2}{r}\right]\psi = E\psi, \quad (1)$$

where $\omega_l = |e|B/2mc$, $L_z = [\boldsymbol{r} \times \boldsymbol{p}]_z$, $\boldsymbol{B} = B\hat{z}$ is a constant, and $\sigma_z$ is the Pauli spinor. Equation (1) is the eigenvalue problem all workers attempt to solve with variational techniques or numerical approximations, once perturbation theory breaks down.

Trial functions, for which all quantities can be determined analytically, can be characterized as being either hydrogenic or pure magnetic. Rajagopal *et al.* used a superposition of $1S$ and $2S$ hydrogenic wave functions;[1] while Cohen *et al.*, Canuto and Kelly, and Yafet *et al.* chose trial functions from the pure magnetic case. The resulting binding energies as a function of magnetic field are shown in figure 1 (indicated by open diamonds, filled circles, open circles, and filled squares, respectively).

## III. AN ATTEMPT AT A SIMPLE TWO-PARAMETER TRIAL FUNCTION

Because in the intermediate field region the atom's symmetry is changing, other workers have had to "superpose" hydrogenic wave functions to have an adequate trial function. However, in the spirit of choosing a simple trial function, the two-parameter form

$$\psi = (\lambda\nu^2/\pi a_0^3)^{1/2} \exp\left[-(\nu^2\rho^2 + \lambda^2 z^2)^{1/2}/a_0\right] \quad (2)$$

seems to be adequate for many purposes. Since $[\nu^2\rho^2 + \lambda^2 z^2]^{1/2} = \nu[\rho^2 + (\lambda/\nu)^2 z^2]^{1/2}$, the parameter $\nu$ corresponds to the "effective charge" in a hydrogenic function (if $\lambda/\nu$ were unity), while the ratio $(\lambda/\nu)$ permits the wave function to be cigar-shaped instead of spherical. Thus, I expected that the form would be somewhat in keeping with the effect of a magnetic field. Using equations (1) and (2), and assuming spin-down, gives, after some rearrangement,

$$\langle H\rangle = \left(\tfrac{2}{3}\nu^2 + \tfrac{1}{3}\lambda^2 - 2\lambda F(\lambda/\nu) + \frac{1}{2}\frac{\chi^2}{\nu^2} - \chi\right)\frac{\hbar^2}{2ma_0^2}, \quad (3)$$

[1] Actually the super position requires an extended calculation with four parameters and one constraint and does not entirely belong in the group of simple analytic wave functions.

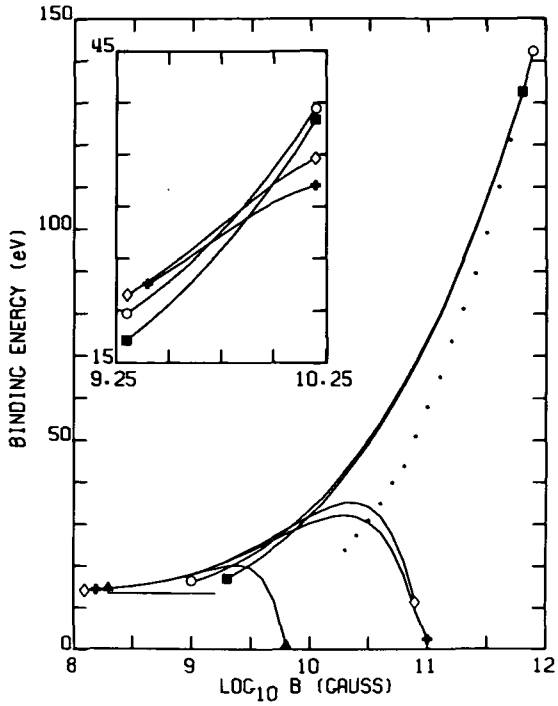



FIG. 1.—The ground state binding energy of hydrogen as a function of the magnetic field $B$ calculated by different workers. The curve indicated by ▲ used perturbation theory, ◇ indicates the Rajagopal *et al.* calculation, + indicates present work for intermediate fields (eq. [5]), ○ indicates the Yafet *et al.* calculation, ■ indicates the Canuto and Kelly calculation, the dotted curve is the Cohen *et al.* calculation (as estimated from their curve), and the solid line indicates 1 Rydberg. The inset, using the same symbols, is a blowup of the transition region.

where

$$\chi = \frac{\hbar\omega_l}{(\hbar^2/2ma_0^2)} = \frac{B(\text{gauss})}{2.350 \times 10^9}$$

and $F$ is given by equation (7). The requirement that $\langle H \rangle$ be a minimum leads to the equations

$$\tfrac{2}{3}\nu - \frac{1}{2}\frac{\chi^2}{\nu^2} = -g_1(\lambda/\nu)\,, \qquad (4a)$$

$$\tfrac{2}{3}\lambda = 2g_2(\lambda/\nu)\,, \qquad (4b)$$

where $g_1$, $g_2$ are given by equation (9). (Note that the coupled equations in (4) can be solved without an iterative numerical minimization technique.) The binding energy curve for this case is also shown in figure 1 (indicated by +). The loss of accuracy is noticeable at high fields, but, at the very least, equation (2) provides an easily calculable first approximation for magnetic fields less than $\sim 5.6 \times 10^9$ gauss.

## IV. COMPARISON OF TRIAL FUNCTIONS

### *a) Resulting Binding-Energy Transition Region*

The various analytic trial functions described earlier have been aimed at either intermediate or intense fields. They tend to behave poorly in the transition region. However, the "prescription" of using equation (2) for magnetic fields less than $5.6 \times 10^9$ gauss and the Yafet *et al.* function for greater fields meets the requirement of a set of simple wave functions which give a relatively smooth transition. Figure 1, particularly the inset portion, shows that this combination provides a much smoother transition from the "hydrogenic" region than the apparent transition shown in the first figure of Rajagopal *et al.* Their transition was uneven because only the binding energy curve of Cohen *et al.* was used for graphical comparison with their work. An even smoother transition using numerical techniques is possible but not of interest here (see § I).

### *b) Resulting Mean Radius*

In addition to increasing the ground-state binding energy, intermediate and intense magnetic fields cause the mean radius, $\langle r \rangle$, to contract. The curves of $\langle r \rangle$ versus field strength are shown in figure 2. The different trial functions yield quite different mean radii, but similar binding energies—a characteristic of most variational calculations (energy differences are "second order"). Again the prescription of § IV*a* provides a reasonable transition.

## V. EXPLICIT CALCULATION OF THE LIMIT AS $B \to \infty$

I will use the Yafet *et al.* trial function to obtain an analytic form for the Hamiltonian. With a slight change in notation, the Yafet *et al.* trial function is

$$\psi = \left(\frac{\lambda\nu^2}{R^3}\right)^{1/2} \pi^{-3/4} \exp\left[-\tfrac{1}{2}(\nu^2\rho^2 + \lambda^2 z^2)/R^2\right] \qquad (5)$$

with $\lambda$, $\nu$ parameters, and $R = (2\hbar c/|B|)^{1/2}$. Using this $\psi$

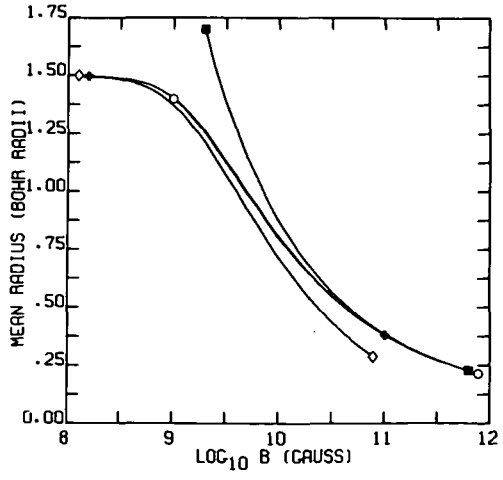



FIG. 2.—The mean radius of hydrogen (in units of the Bohr radius) as a function of the magnetic field $B$ calculated using different variational functions. The curve indicated by ◇ used the Rajagopal *et al.* function, + used the present work's function for intermediate fields, ○ used the Yafet *et al.* function, ■ used the Canuto and Kelly function.

in the Hamiltonian (1) and assuming spin down, gives

$$\langle H\rangle = \left[\left(\nu - \frac{1}{\nu}\right)^2 + \tfrac{1}{2}\lambda^2 - \frac{4}{\pi^{1/2}}\frac{R}{a_0}F\left(\frac{\lambda}{\nu}\right)\right] \times \left(\frac{a_0}{R}\right)^2\left(\frac{\hbar^2}{2ma_0^2}\right), \quad (6)$$

where $F(\lambda/\nu)$ comes from the Coulomb term and is the continuous function

$$F\left(\frac{\lambda}{\nu}\right) = \left[1 - \left(\frac{\lambda}{\nu}\right)^2\right]^{-1/2}\tan^{-1}\left[1 - \left(\frac{\lambda}{\nu}\right)^2\right]^{1/2}, \quad (\lambda/\nu) < 1,$$
$$= 1, \quad \lambda = 1, \quad (7)$$
$$= \left[\left(\frac{\lambda}{\nu}\right)^2 - 1\right]^{-1/2}\tan^{-1}\left[\left(\frac{\lambda}{\nu}\right)^2 - 1\right]^{1/2}, \quad (\lambda/\nu) > 1.$$

The parameters $\lambda$ and $\nu$ are determined by the equations

$$2(\nu - 1/\nu^3)(a_0/R) = 4\pi^{-1/2}g_1(\lambda/\nu), \quad (8a)$$
$$\lambda(a_0/R) = 4\pi^{-1/2}g_2(\lambda/\nu), \quad (8b)$$

where

$$g_1(x) = -[x/(1 - x^2)][x^2F(x) - 1] \quad (9a)$$

and

$$g_2(x) = [1/(1 - x^2)][F(x) - 1]. \quad (9b)$$

The calculation of the limit is done by computing the limit of equation (6) as $\lambda/\nu \to 0$, and also demonstrating that $\lambda/\nu \to 0$ implies $B \to \infty$. Let $\lambda/\nu = \epsilon \ll 1$. Then equation (7)–(9) to lowest order give

$$\nu = 1 - \tfrac{1}{8}\epsilon^2/\ln\epsilon. \quad (10)$$

Since equation (10) implies $\lambda = \epsilon$, equation (8b) can be used to show that $a_0/R$ diverges. That is,

$$\frac{a_0}{R} = \frac{4}{\pi^{1/2}}\frac{\ln(1/\epsilon)}{\epsilon} \quad (11)$$

Since $R \propto B^{-1/2}$, equation (11) verifies the claim that $\lambda/\nu = \epsilon \to 0$ implies $B \to \infty$. With $\lambda$, $\nu$, and $a_0/R$ expressed in terms of $\epsilon$, the ground-state energy (eq. [6]) becomes (to the lowest order)

$$\langle H\rangle = -\frac{8}{\pi}\left[\ln\left(\frac{1}{\epsilon}\right)\right]^2\frac{\hbar^2}{2ma_0^2}. \quad (12)$$

But if $\epsilon$ is truly infinitesimal, then $\ln(1/\epsilon) \gg \ln[\ln(1/\epsilon)]$, and equation (11) can be used to rewrite equation (12) in the more physical form

$$\langle H\rangle = -\frac{2}{\pi}\left[2\ln\left(\frac{a_0}{R}\right)\right]^2\frac{\hbar^2}{2ma_0^2}, \quad (13)$$

Equation (13) is to be compared with the analytical results of Haines and Roberts (1969) for the nonrelativistic one-dimensional hydrogen atom in a truncated Coulomb potential. With the potential given by

$$V(x) = -e^2/(|x| + a), \quad a > 0,$$

they found that the ground-state energy depended sensitively on the truncation parameter $a$. Assuming $a_0 \gg a$, they found that the ground state was given by

$$\langle H\rangle_{1\,\mathrm{dim}} = -\left[2\ln\left(\frac{a_0}{a}\right)\right]^2\frac{\hbar^2}{2ma_0^2}$$

Thus, even though the gyroradius $R$ is the natural truncation length for this problem, the intense magnetic field limit of this trial function does not cause a complete collapse of the perpendicular degrees of freedom. Instead, the Yafet *et al.* trial function gives

$$H = 2\pi^{-1}H_{1\,\mathrm{dim}}, \quad as \quad B \to \infty.$$

## VI. DISCUSSION

Somewhat surprisingly, accretion models for regularly pulsing X-ray sources (Lamb, Pethick, and Pines 1973; Pringle and Rees 1972; Davidson and Ostriker 1973) may need to incorporate a contribution to the X-ray opacity from the bound-bound and bound-free transitions of neutral hydrogen. Assuming that the accretion columns near the magnetic poles are in thermal equilibrium and using the Lamb *et al.* model for density and temperature ($\rho \sim 700\ \mathrm{g\ cm^{-3}}$, $T \sim 10^8\ ^\circ\mathrm{K}$), it can be inferred from the Saha equation that $\sim 13$ percent of the hydrogen still remains neutral. The temperature is so high that the fraction of neutrals does not depend very sensitively on the magnetic field ($\exp[-\chi/T] \approx 1$). However, the thresholds for transitions from bound states rise to a few hundred electron volts in fields greater than about $10^{12}$ gauss. Hence, the transport of very soft X-rays could be affected by the presence of 13 percent neutral hydrogen.

## VII. CONCLUSIONS

1. As discussed in § IV*a*, the simple "prescription" of using equation (2) for intermediate fields less than $5.6 \times 10^9$ gauss and the Yafet *et al.* function (eq. [5]) for greater fields provides a relatively smooth transition for the ground-state binding energy. The uneven transition shown by Rajagopal *et al.* arises because only the Cohen *et al.* curve appeared in their figure.

2. The Yafet *et al.* function yields a limiting ground-state energy of

$$\langle H\rangle = -\frac{2}{\pi}\left[2\ln\left(\frac{a_0}{R}\right)\right]^2\left(\frac{\hbar^2}{2ma_0^2}\right)$$

for arbitrarily large fields. As discussed in § V, this logarithmic divergence for $\langle H\rangle$ is analogous to the behavior of the one-dimensional hydrogen atom, but has a different coefficient.

I wish to thank Drs. P. B. Price and R. Cowsik for suggesting the problem and giving much encouragement along the way. This research was supported by NASA grant NGR 05-003-376.

## REFERENCES


Baldereschi, A., and Bassani, F. 1970, *Tenth International Conference on the Physics of Semiconductors* (Cambridge, Mass.: Atomic Energy Commission), 191.

Cabib, D., Fabri, E., and Fiorio, G. 1971, *Solid State Comm.*, **9**, 1517.

———. 1972, *Nuovo Cimento*, **10B**, 185.

Callaway, Joseph. 1972, *Phys. Letters*, **40A**, 331.

Canuto, V., and Kelly, D. C. 1972, *Ap. and Space Sci.*, **17**, 27.

Cohen, R., Lodenquai, J., and Ruderman, M. 1970, *Phys. Rev. Letters*, **25**, 467.

Davidson, K., and Ostriker, J. P. 1973, *Ap. J.*, **179**, 585.

Haines, L. K., and Roberts, D. H. 1969, *Amer. J. Phys.*, **37**, 11.

Kemp, J. C. 1970, *Ap. J.*, **162**, 169.

Kemp, J. C., Swedlund, J. B., Landstreet, J. D., and Angel, J. R. P. 1970, *Ap. J. (Letters)*, **161**, L77.

Lamb, F. K., Pethick, C. J., and Pines, D. 1973, *Ap. J.*, **184**, 271.

Larsen, D. M. 1968, *J. Phys. Chem. Solids*, **29**, 271.

Pokatilov, E. P., and Rusanov, M. M. 1969, *Soviet Phys. Solid State*, **10**, 2458.

Pringle, J. E., and Rees, M. J. 1972, *Astr. and Ap.*, **21**, 1, 1–9.

Rajagopal, A. K., Chanmugam, G., O'Connell, R. F., and Surmelian, G. L. 1972, *Ap. J.*, **177**, 713.

Yafet, Y., Keyes, R. W., and Adams, E. N. 1956, *J. Phys. Chem. Solids*, **1**, 137.

THE ASTROPHYSICAL JOURNAL, 187:L19–L21, 1974 January 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

# THE METHANOL SOURCE IN ORION AT 1.2 CENTIMETERS

M. F. CHUI, A. C. CHEUNG, D. MATSAKIS*, AND C. H. TOWNES†
Department of Physics, University of California, Berkeley 94720 (415)642-1128

AND

A. G. CARDIASMENOS‡
Department of Electrical and Computer Engineering, University of Massachusetts, Amherst 01002
*Received 1973 November 5*

## ABSTRACT

Methanol line radiation from Orion near 25 GHz has been reexamined and found to be much stronger than expected from earlier reports. The line widths are also narrower than other nonmasing molecular radiation. These results indicate either that there is more than one small source of methanol radiation in Orion, or that these transitions are inverted and varying with time.

*Subject headings:* interstellar matter — molecules

Barrett, Schwartz, and Waters (1971) have reported emission from the $6_2 \rightarrow 6_1$ ($E_1$) and $7_2 \rightarrow 7_1$ ($E_1$) transitions of $CH_3OH$ in the direction of the Orion Nebula. We recently remeasured these transitions in the process of checking system performance of a newly installed maser preamplifier, and found that the integrated flux densities are about 4 times stronger than is expected from the previous results.

The observations were made in 1973 between January and July with the Hat Creek 20-foot (6-m) telescope equipped with a maser preamplifier (Yngvesson *et al.* 1974). At 25.1 GHz, the total system noise is typically 200° K on zenith. The receiver was switched between two feed horns each 7′.8 in size and separated by 22′. The antenna temperatures were calibrated by measuring the Orion continuum, which was in turn calibrated against a liquid nitrogen cold load. A bank of 64 filters, each 15 kHz wide and separated by 15 kHz from contiguous ones, was used for these measurements.

The observed spectra of the $J = 6$ line at 25.01814 GHz and the $J = 7$ line at 25.12488 are presented in figure 1. The $J = 6$ line has a peak antenna temperature of 1.0° ± 0.07° K (200 f.u. for 50% aperture efficiency at a velocity (with respect to the local standard of rest) of +8.3 km s$^{-1}$ and a line width of 53 ± 8 kHz (0.64 ± 0.1 km s$^{-1}$). The $J = 7$ line has a peak $T_A$ of 0.7° ± 0.06° K (140 f.u.) at the same LSR velocity and a line width of 93 ± 8 kHz (1.1 ± 0.1 km s$^{-1}$). A nominal atmospheric absorption of 10 percent has been taken into account. The split of the $J = 7$ line appears to be real, and occurs at the velocity at which the $J = 6$ line peaks.

* NDEA Fellow.

† Work supported in part by Office of Naval Research Contract N00014-69-A-0200-1003 and National Science Foundation Grant GP30424X.

‡ Work supported in part by National Science Foundation Grant GP38593X.

The $CH_3OH$ source measured by Barrett *et al.* (1971) with the 120-foot (37-m) Haystack antenna was reported to be less than 1 arc minute in size. The $J = 6$ and $J = 7$ lines had antenna temperatures of 1.7° and 1.5° K, respectively, equivalent for small sources to flux densities of 26 and 23 f.u., which are factors of 8 and 6 less than those measured on the 20-foot. If a weak component of the expected intensity and width of 150 kHz is subtracted from each observed spectrum,

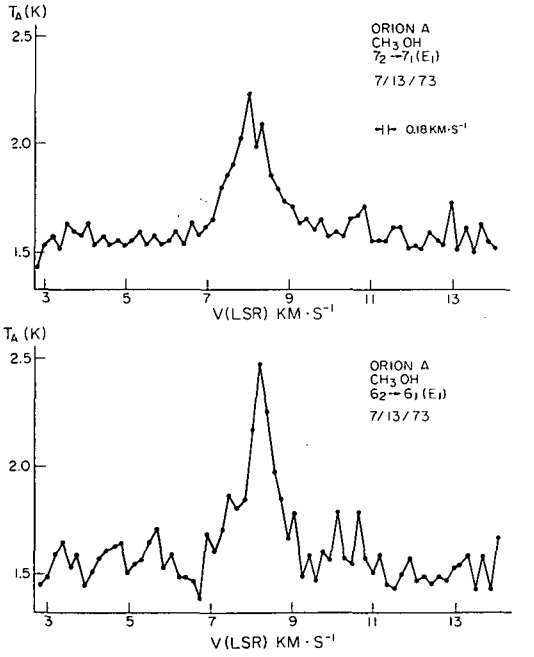



FIG. 1.—$CH_3OH$ emission spectra in the direction of Orion *A* ($\alpha_{1950} = 5^h32^m47^s$, $\delta_{1950} = -5°\ 24'.4$). (*a*) $7_2 \rightarrow 7_1(E_1)$ transition, (*b*) $6_2 \rightarrow 6_1(E_1)$ transition.

the lines are not appreciably changed in intensity or width.

The line widths are narrower than those of any other molecules observed in the Orion molecular cloud except for the OH and $H_2O$ maser sources. This places severe restrictions on the size of the source of additional radiation, which will now be discussed.

Since our beamwidth of 7'.8 covers a larger area than the region of about 4' mapped by Barrett *et al.*, it is possible that the line emission comes from a rather broad region. If the observed excess emission comes from a uniform region as large as 8', then it would produce a temperature $T_A$ of 0.3° K on the 120-foot antenna which would be within the noise level of the system used by Barrett *et al.*, and would be undetected. In this calculation, the beam efficiencies for the 120- and 20-foot antennas are taken to be 20 and 70 percent, respectively.

Other molecules such as CO (Liszt *et al.* 1974), $H_2CO$ (Harvey *et al.* 1974), and $NH_3$ observed in this direction show a LSR velocity spread of about 2 km $s^{-1}$ over a few arc minutes extent. Irrespective of its local Doppler line width, therefore, $CH_3OH$ radiation from such a large cloud should have a line width of at least 2 km $s^{-1}$. Assuming an intensity as large as the 3 $\sigma$ non-detection limit of Barrett *et al.*, the source size would need to be about 6'.5 in order to provide the presently observed intensity. This size, however, is still too large to yield the narrow line widths (0.7 km $s^{-1}$) observed. Therefore, the source of additional radiation cannot be large and uniform. Furthermore, if it is a source which was not observed by Barrett *et al.*, it must be displaced from their source by at least 2', since this is the distance over which they searched for methanol emission.

Kutner *et al.*, (1973) detected $CH_3OH$ at $\lambda = 2$ mm at 7.5 km $s^{-1}$ in the direction of the Kleinmann-Low nebula (central position), which is where Barrett *et al.* observed $CH_3OH$ radiation at 8.5 km $s^{-1}$, and another source displaced 2' south of this position. This southern source reported by Kutner *et al.* is a candidate for the additional flux at 25 GHz. The apparent difference in LSR velocity between the millimeter and centimeter lines at the central position is within the uncertainties of measurement (Kutner, private communication). An upper limit may be placed on the size of the postulated source by assuming the velocity gradient of the 2-cm $H_2CO$ emission (Harvey *et al.* 1974) in the Orion molecular cloud. For a line width of 0.7 km $s^{-1}$, the source is spread over at most 2'. This sets a lower limit of 30° K on the brightness temperature of the $J = 6$ line.

The ratio of the integrated line intensities of the $J = 6$ and $J = 7$ lines is not appreciably different from that observed by Barrett *et al.* Therefore, if one assumes local thermodynamic equilibrium, approximately the same kinetic temperature of 90° K is obtained. If the source size is 2', the column density of $CH_3OH$ is about $10^{17}$ $cm^{-2}$. However, since the line width of the $J = 7$ transition appears to be wider than that of the $J = 6$, it suggests that the $J = 7$ radiation originates from a more turbulent and hotter region, possibly embedded inside the region which provides most of the $J = 6$ radiation.

Finally, the most straightforward interpretation of the observations is that there have been time variations in the source characteristics in the two years between the original measurements of Barrett *et al.* and the present ones. There have been conjectures that these particular transitions are nonthermal and possibly

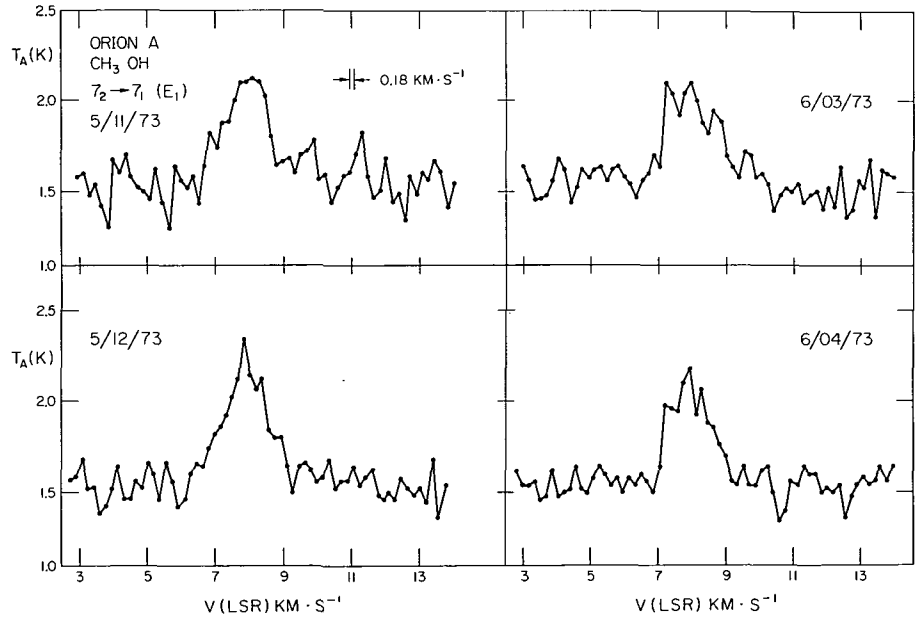



FIG. 2.—$CH_3OH$ $7_2 \rightarrow 7_1(E_1)$ emission spectra in Orion A showing possible changes in line shape with time

inverted (Barrett *et al.* 1971; Turner, Gordon, and Wrixon 1972; Gottlieb *et al.* 1973), which would make time variations reasonable. The narrower-than-usual line widths also give some indication of maser action in the source. However, the $J = 7$ line has been monitored for about 3 months. Figures 1*a* and 2 show $J = 7$ spectra taken over several days during this period. There is some indication from these spectra that the line changed in shape between May and June, but noise fluctuations in the measurements make this uncertain. Furthermore, the line-center velocity has not changed, nor has the integrated intensity changed by more than 30 percent, approximately the accuracy of present measurements. In addition, if the $J = 6$ and $J = 7$ lines have varied over the last two years, their integrated intensities have remained in the same ratio.

The above considerations show that if the $CH_3OH$ source in Orion is not a maser then $CH_3OH$ occurs in at least two regions, separated by at least 2′. In the latter case, the noncentral component would be more intense than the central source by about a factor of 3. Additional study with a narrow beam should discriminate between these two cases.

We thank Professors W. J. Welch and K. S. Yngvesson for advice and help with the design and installation of the maser receiver, Professor A. H. Barrett and Dr. P. Schwartz for discussions of their results, and the staff of the Hat Creek Observatory for assistance with observations.

## REFERENCES

Barrett, A. H., Schwartz, P. R., and Waters, J. W. 1971, *Ap. J.* (*Letters*), **168**, L101.

Gottlieb, C. A., Palmer, P., Rickard, L. J. M., and Zuckerman, B. 1973, *Ap. J.*, **182**, 699.

Harvey, P. M., Evans, N. J., II, Werner, M. W., Elias, J., Gatley, I., Morris, G., Sato, G., Zuckerman, B., and Litvak, M. M. 1974, in preparation.

Kutner, M. L., Thaddeus, P., Penzias, A. A., Wilson, R. W., and Jefferts, K. B. 1973, *Ap. J.* (*Letters*), **183**, L27.

Liszt, H. S., Wilson, R. W., Penzias, A. A., Jefferts, K. B., and Solomon, P. M. 1974, *Ap. J.*, in press.

Turner, B. E., Gordon, M. A., and Wrixon, G. T. 1972, *Ap. J.*, **177**, 609.

Yngvesson, K. S., Townes, C. H., Cheung. A. C., Chui, M. F., Cardiasmenos, A. G., and Wang, S. Y. 1974, in preparation.

THE ASTROPHYSICAL JOURNAL, 187:L111–L112, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

A74-31960

# DETECTION OF $^{17}O$ IN IRC+10216*

D. M. RANK
Lick Observatory, Board of Studies in Astronomy and Astrophysics,
University of California, Santa Cruz 95060

AND

T. R. GEBALLE AND E. R. WOLLMAN
Department of Physics, University of California, Berkeley 94720
*Received 1973 October 24; revised 1973 December 3*

## ABSTRACT

Lines of the fundamental vibration-rotation band of $^{12}C^{17}O$ have been observed near 4.7 $\mu$ in IRC+10216. Lines of $^{12}C^{18}O$ were not detected, indicating a lower limit of 5 for the abundance ratio $^{17}O/^{18}O$. A lower limit of 200 and a probable value of 400 is obtained for the ratio $^{16}O/^{17}O$. The presence of $^{12}C^{17}O$ is also indicated in $\alpha$ Her and $\alpha$ Sco.

*Subject headings:* abundances — infrared — spectra, molecular

In an earlier paper (Geballe, Wollman, and Rank 1973, hereafter denoted GWR), it was suggested that an absorption feature at 2142.0 $cm^{-1}$ in the bright infrared source IRC+10216 could be identified with the 1–0 *R*6 transition of $^{12}C^{17}O$. Observations of this feature have since been repeated. In addition a second line of $^{12}C^{17}O$ has been observed in IRC+10216, confirming the identification of this isotopic species.

The spectra shown in figures 1 and 2 were obtained at the Lick Observatory's 120-inch (3-m) telescope with a tandem scanning Fabry-Perot interferometer operated at a resolution of 0.15 $cm^{-1}$. Figure 1 should be compared with figure 1 of GWR. The two spectra were taken approximately 6 months apart. All lines inherent to IRC+10216 have been shifted to lower frequency by about 0.4 $cm^{-1}$ in the present spectrum due to the change in the radial velocity of IRC+10216 relative to the Earth. In particular, the absorption at the frequency of the 1–0 *R*6 line of $^{12}C^{17}O$ is now well isolated from the terrestrial $H_2O$ absorption and, in view of its large signal-to-noise ratio, can no longer be regarded as tentative. Figure 2 shows a spectral region near that of figure 1. In it, the $^{13}C^{16}O$ line at $\sim$2134.65 $cm^{-1}$ has a shoulder on its high-frequency side whose frequency agrees with that of the 1–0 *R*4 transition of $^{12}C^{17}O$ and whose strength is consistent with the previously mentioned *R*6 line of $^{12}C^{17}O$. A solar spectrum taken at the time showed no terrestrial absorptions other than the strong feature near 2135.5 $cm^{-1}$. Thus, figure 1 shows detection of an isolated line corresponding to $^{12}C^{17}O$ and figure 2 provides confirmatory evidence for $^{12}C^{17}O$ in IRC+10216.

The solid lines superposed on the data points are computer-generated synthetic spectra corresponding to absorption by an isothermal shell with a Gaussian line profile. The computed stellar spectrum was convolved with the instrumental profile, and the parameters of the absorbing layer were adjusted to produce a fit to the data. These parameters are listed in table 1. More complete discussions of the mathematics of the fit and the details of the structure of IRC+10216 can be found in Geballe, Wollman, and Rank (1972) and in GWR.[1]

The positions of the transitions of the fundamental vibration-rotation band of $^{12}C^{18}O$ are also shown in figures 1 and 2. The absence of absorption at these frequencies yields a lower limit of 5 for the abundance ratio $^{17}O/^{18}O$. Thus, $^{17}O/^{18}O$ in IRC+10216 is at least 25 times greater than the terrestrial value and the recently reported solar value (Hall 1973). The $^{16}O/^{17}O$ ratio of 400 is based on a theoretical best fit to the data which minimizes the $^{12}C^{16}O$ column density in the shell around IRC+10216. Since the $^{12}C^{16}O$ lines are saturated, an acceptable synthetic spectrum can be computed with nearly an order of magnitude more $^{12}C^{16}O$ and, consequently, an $^{16}O/^{17}O$ ratio which is essentially terrestrial. However, a standard C/H ratio in the shell would then imply a shell mass for IRC+10216 which is a few tenths of a solar mass. The large mass and associated optical extinction of the shell, coupled with the high ratio of $^{16}O/^{18}O$ ($>10^4$) required by such an interpretation of the data seem to us to be unrealistic. Consequently, it appears that the $^{16}O/^{17}O$ ratio is substantially lower than its terrestrial value.

A discussion of the production and surface enrichment of $^{17}O$ has been given by Wollman (1973). A low $^{16}O/^{17}O$ ratio in the material ejected by IRC+10216 requires post–main-sequence convection down to the region where partial NO processing has occurred. This is consistent with the post–main-sequence CN processing of the convective envelope required to produce the

* Work partially supported by NASA grants NGL 05-003-272, NGR 05-003-452, and NGR 05-061-015.

[1] Due to an error in calculation, the ages of the shells stated in GWR should be increased by approximately a factor of 4, and the column density of $H_2$ in the shell around the star reduced by the same factor. Accordingly, the total mass of the 2″ shell should be increased by the square of this factor.

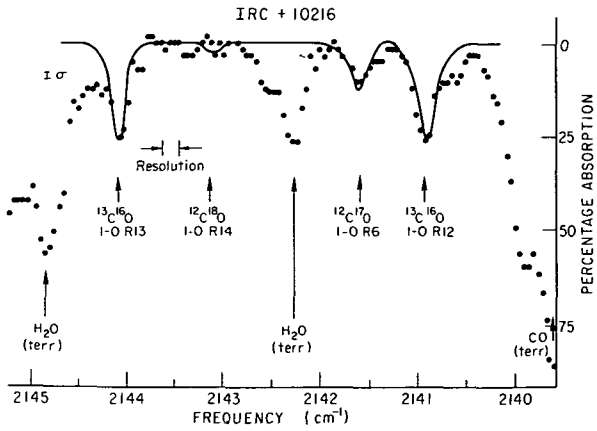



FIG. 1.—Infrared spectrum of IRC+10216 near 2142 cm$^{-1}$. *Dots*, observed spectra; $\sigma$ is the probable error for a single data point. The solid line is a synthetic spectrum computed using the values for the parameters listed in table 1. The positions of terrestrial lines and possibly detectable lines in IRC+10216 are indicated by arrows.

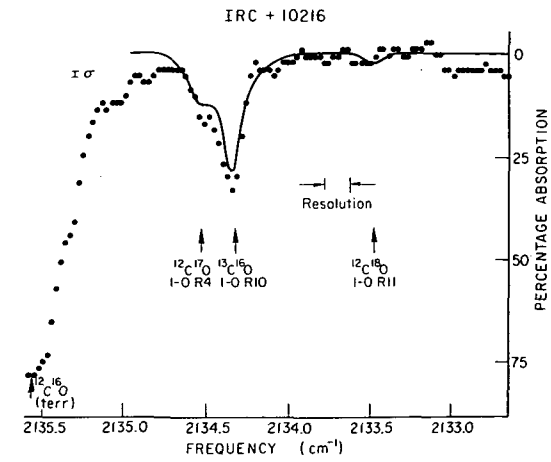



FIG. 2.—Infrared spectrum of IRC+10216 near 2134 cm$^{-1}$. Note that the $^{12}C^{17}O$ and $^{13}C^{16}O$ lines near 2134.5 cm$^{-1}$ are on the edge of a broad terrestrial $^{12}C^{16}O$ absorption. Hence the stellar spectrum falls below the synthetic spectrum in this region.

TABLE 1

PARAMETERS FOR SIMULATION OF OBSERVED IRC+10216 SPECTRA

| Assumed $T_{eff}$(° K) | $T_R$(° K) | $T_V$(° K) | $N(^{12}C^{16}O)$ (cm$^{-2}$) | Line-Width Parameter $\sigma$(cm$^{-1}$) | $^{12}C/^{13}C$ | $^{17}O/^{18}O$ | $^{16}O/^{17}O$ |
|---|---|---|---|---|---|---|---|
| 650 | 400 | <300 | $10^{19.7}$ | 0.009 | 8* | >5 | 400† |

* Value obtained from measurements by Wilson *et al.* (1971).
† Value uncertain due to saturation effects.

low value of $^{12}C/^{13}C$ observed in IRC+10216 (see, e.g., Wilson *et al.* 1973, and GWR).

In addition to the spectra discussed above, we have recently observed the two late-type stars $\alpha$ Her and $\alpha$ Sco in spectral regions near 4.7 $\mu$ for the purpose of identifying lines of $^{12}C^{17}O$ and $^{12}C^{18}O$. Due to improvements in the Fabry-Perot spectrometer these spectra are of higher quality than those reported by Geballe *et al.* (1972) for these two stars. The profusion of lines of $^{12}C^{16}O$ and $^{13}C^{16}O$ from the first several vibrational levels makes identification of lines of rarer isotopic species very difficult. However, in both of these stars there is evidence for the presence of $^{12}C^{17}O$, indicating an $^{16}O/^{17}O$ abundance ratio of 200 with an estimated error of a factor of 3.[2] The presence of $^{12}C^{18}O$, which had been tentatively identified in $\alpha$ Her by Geballe *et al.* (1972), is uncertain, and its abundance is much less than was reported in that paper. The present data do not require a substantial deviation from the terrestrial $^{16}O/^{18}O$ ratio in $\alpha$ Her or $\alpha$ Sco.

The authors are grateful to Professor C. H. Townes for several helpful discussions. We also thank A. L. Isaac for assistance with the observations and data reduction.

REFERENCES

Geballe, T. R., Wollman, E. R., and Rank, D. M. 1972, *Ap. J. (Letters)*, **177**, L27.
———. 1973, *Ap. J.*, **183**, 499.
Hall, D. N. B. 1973, *Ap. J.*, **182**, 977.
Maillard, J. M. 1973, paper given at the IAU 15th General Assembly.
Wilson, R. W., Solomon, P. M., Penzias, A. A., and Jefferts, K. B. 1971, *Ap. J. (Letters)*, **169**, L35.
Wilson, W. J., Schwartz, P. R., and Epstein, E. E. 1973, *Ap. J.*, **183**, 871.
Wollman, E. R. 1973, *Ap. J.*, **184**, 773.

[2] Maillard (1973) has also recently reported the enrichment of $^{17}O$ in $\alpha$ Her from observations of the overtone bands of CO.

THE ASTROPHYSICAL JOURNAL, 187:L111–L112, 1974 February 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

# DETECTION OF $^{17}O$ IN IRC+10216*

D. M. RANK
Lick Observatory, Board of Studies in Astronomy and Astrophysics,
University of California, Santa Cruz 95060

AND

T. R. GEBALLE AND E. R. WOLLMAN
Department of Physics, University of California, Berkeley 94720
*Received 1973 October 24; revised 1973 December 3*

## ABSTRACT

Lines of the fundamental vibration-rotation band of $^{12}C^{17}O$ have been observed near 4.7 $\mu$ in IRC+10216. Lines of $^{12}C^{18}O$ were not detected, indicating a lower limit of 5 for the abundance ratio $^{17}O/^{18}O$. A lower limit of 200 and a probable value of 400 is obtained for the ratio $^{16}O/^{17}O$. The presence of $^{12}C^{17}O$ is also indicated in $\alpha$ Her and $\alpha$ Sco.

*Subject headings:* abundances — infrared — spectra, molecular

In an earlier paper (Geballe, Wollman, and Rank 1973, hereafter denoted GWR), it was suggested that an absorption feature at 2142.0 $cm^{-1}$ in the bright infrared source IRC+10216 could be identified with the 1–0 *R*6 transition of $^{12}C^{17}O$. Observations of this feature have since been repeated. In addition a second line of $^{12}C^{17}O$ has been observed in IRC+10216, confirming the identification of this isotopic species.

The spectra shown in figures 1 and 2 were obtained at the Lick Observatory's 120-inch (3-m) telescope with a tandem scanning Fabry-Perot interferometer operated at a resolution of 0.15 $cm^{-1}$. Figure 1 should be compared with figure 1 of GWR. The two spectra were taken approximately 6 months apart. All lines inherent to IRC+10216 have been shifted to lower frequency by about 0.4 $cm^{-1}$ in the present spectrum due to the change in the radial velocity of IRC+10216 relative to the Earth. In particular, the absorption at the frequency of the 1–0 *R*6 line of $^{12}C^{17}O$ is now well isolated from the terrestrial $H_2O$ absorption and, in view of its large signal-to-noise ratio, can no longer be regarded as tentative. Figure 2 shows a spectral region near that of figure 1. In it, the $^{13}C^{16}O$ line at $\sim$2134.65 $cm^{-1}$ has a shoulder on its high-frequency side whose frequency agrees with that of the 1–0 *R*4 transition of $^{12}C^{17}O$ and whose strength is consistent with the previously mentioned *R*6 line of $^{12}C^{17}O$. A solar spectrum taken at the time showed no terrestrial absorptions other than the strong feature near 2135.5 $cm^{-1}$. Thus, figure 1 shows detection of an isolated line corresponding to $^{12}C^{17}O$ and figure 2 provides confirmatory evidence for $^{12}C^{17}O$ in IRC+10216.

The solid lines superposed on the data points are computer-generated synthetic spectra corresponding to absorption by an isothermal shell with a Gaussian line profile. The computed stellar spectrum was convolved with the instrumental profile, and the parameters of the absorbing layer were adjusted to produce a fit to the data. These parameters are listed in table 1. More complete discussions of the mathematics of the fit and the details of the structure of IRC+10216 can be found in Geballe, Wollman, and Rank (1972) and in GWR.[1]

The positions of the transitions of the fundamental vibration-rotation band of $^{12}C^{18}O$ are also shown in figures 1 and 2. The absence of absorption at these frequencies yields a lower limit of 5 for the abundance ratio $^{17}O/^{18}O$. Thus, $^{17}O/^{18}O$ in IRC+10216 is at least 25 times greater than the terrestrial value and the recently reported solar value (Hall 1973). The $^{16}O/^{17}O$ ratio of 400 is based on a theoretical best fit to the data which minimizes the $^{12}C^{16}O$ column density in the shell around IRC+10216. Since the $^{12}C^{16}O$ lines are saturated, an acceptable synthetic spectrum can be computed with nearly an order of magnitude more $^{12}C^{16}O$ and, consequently, an $^{16}O/^{17}O$ ratio which is essentially terrestrial. However, a standard C/H ratio in the shell would then imply a shell mass for IRC+10216 which is a few tenths of a solar mass. The large mass and associated optical extinction of the shell, coupled with the high ratio of $^{16}O/^{18}O$ ($>10^4$) required by such an interpretation of the data seem to us to be unrealistic. Consequently, it appears that the $^{16}O/^{17}O$ ratio is substantially lower than its terrestrial value.

A discussion of the production and surface enrichment of $^{17}O$ has been given by Wollman (1973). A low $^{16}O/^{17}O$ ratio in the material ejected by IRC+10216 requires post–main-sequence convection down to the region where partial NO processing has occurred. This is consistent with the post–main-sequence CN processing of the convective envelope required to produce the

* Work partially supported by NASA grants NGL 05-003-272, NGR 05-003-452, and NGR 05-061-015.

[1] Due to an error in calculation, the ages of the shells stated in GWR should be increased by approximately a factor of 4, and the column density of $H_2$ in the shell around the star reduced by the same factor. Accordingly, the total mass of the 2″ shell should be increased by the square of this factor.

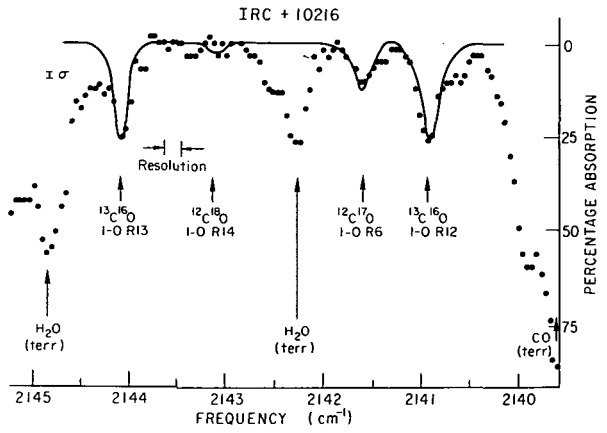



FIG. 1.—Infrared spectrum of IRC+10216 near 2142 cm$^{-1}$. *Dots*, observed spectra; $\sigma$ is the probable error for a single data point. The solid line is a synthetic spectrum computed using the values for the parameters listed in table 1. The positions of terrestrial lines and possibly detectable lines in IRC+10216 are indicated by arrows.

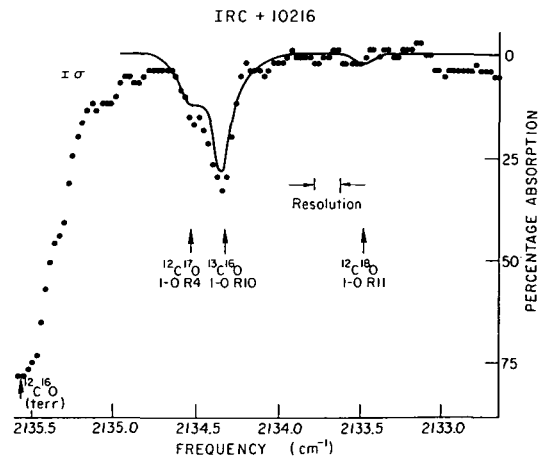



FIG. 2.—Infrared spectrum of IRC+10216 near 2134 cm$^{-1}$. Note that the $^{12}C^{17}O$ and $^{13}C^{16}O$ lines near 2134.5 cm$^{-1}$ are on the edge of a broad terrestrial $^{12}C^{16}O$ absorption. Hence the stellar spectrum falls below the synthetic spectrum in this region.

TABLE 1

PARAMETERS FOR SIMULATION OF OBSERVED IRC+10216 SPECTRA

| Assumed $T_{eff}$(° K) | $T_R$(° K) | $T_V$(° K) | $N(^{12}C^{16}O)$ (cm$^{-2}$) | Line-Width Parameter $\sigma$(cm$^{-1}$) | $^{12}C/^{13}C$ | $^{17}O/^{18}O$ | $^{16}O/^{17}O$ |
|---|---|---|---|---|---|---|---|
| 650 | 400 | <300 | $10^{19.7}$ | 0.009 | 8* | >5 | 400† |

* Value obtained from measurements by Wilson *et al.* (1971).
† Value uncertain due to saturation effects.

low value of $^{12}C/^{13}C$ observed in IRC+10216 (see, e.g., Wilson *et al.* 1973, and GWR).

In addition to the spectra discussed above, we have recently observed the two late-type stars $\alpha$ Her and $\alpha$ Sco in spectral regions near 4.7 $\mu$ for the purpose of identifying lines of $^{12}C^{17}O$ and $^{12}C^{18}O$. Due to improvements in the Fabry-Perot spectrometer these spectra are of higher quality than those reported by Geballe *et al.* (1972) for these two stars. The profusion of lines of $^{12}C^{16}O$ and $^{13}C^{16}O$ from the first several vibrational levels makes identification of lines of rarer isotopic species very difficult. However, in both of these stars there is evidence for the presence of $^{12}C^{17}O$, indicating an $^{16}O/^{17}O$ abundance ratio of 200 with an estimated error of a factor of 3.[2] The presence of $^{12}C^{18}O$, which had been tentatively identified in $\alpha$ Her by Geballe *et al.* (1972), is uncertain, and its abundance is much less than was reported in that paper. The present data do not require a substantial deviation from the terrestrial $^{16}O/^{18}O$ ratio in $\alpha$ Her or $\alpha$ Sco.

The authors are grateful to Professor C. H. Townes for several helpful discussions. We also thank A. L. Isaac for assistance with the observations and data reduction.

REFERENCES

Geballe, T. R., Wollman, E. R., and Rank, D. M. 1972, *Ap. J.* (*Letters*), **177**, L27.
———. 1973, *Ap. J.*, **183**, 499.
Hall, D. N. B. 1973, *Ap. J.*, **182**, 977.
Maillard, J. M. 1973, paper given at the IAU 15th General Assembly.
Wilson, R. W., Solomon, P. M., Penzias, A. A., and Jefferts, K. B. 1971, *Ap. J.* (*Letters*), **169**, L35.
Wilson, W. J., Schwartz, P. R., and Epstein, E. E. 1973, *Ap. J.*, **183**, 871.
Wollman, E. R. 1973, *Ap. J.*, **184**, 773.

[2] Maillard (1973) has also recently reported the enrichment of $^{17}O$ in $\alpha$ Her from observations of the overtone bands of CO.

115947